

Argumentation Semantics for Prioritised Default Logic

Anthony P. Young*, Sanjay Modgil, Odinaldo Rodrigues

1st July 2015

Abstract

We endow prioritised default logic (PDL) with argumentation semantics using the ASPIC⁺ framework for structured argumentation, and prove that the conclusions of the justified arguments are exactly the prioritised default extensions. Argumentation semantics for PDL will allow for the application of argument game proof theories to the process of inference in PDL, making the reasons for accepting a conclusion transparent and the inference process more intuitive. This also opens up the possibility for argumentation-based distributed reasoning and communication amongst agents with PDL representations of mental attitudes.

Contents

1	Introduction	2
2	Background	4
2.1	Dung’s Abstract Argumentation Theory	4
2.2	The ASPIC ⁺ Framework	4
2.2.1	Construction of Arguments and Knowledge Bases	5
2.2.2	Properties of Arguments	6
2.2.3	Attacks	7
2.2.4	Preferences and Defeats	7
2.2.5	Applying Dung’s Semantics	9
2.2.6	Conditions for Normative Rationality	9
2.3	Prioritised Default Logic	9
2.3.1	First Order Logic	9
2.3.2	Normal Defaults	9
2.3.3	Prioritised Default Theories and Extensions	10

***Corresponding author:** Peter Young, Department of Informatics, King’s College London, anthony.p.young@kcl.ac.uk

3	The Corrected ASPIC⁺ Preferences	11
3.1	Overview of the Problem	11
3.2	The Elitist Set Comparison Relation is not Reasonable Inducing	11
3.3	The Strict Elitist Set Comparison is Reasonable Inducing	13
3.4	Summary	16
4	From ASPIC⁺ to Prioritised Default Logic	16
4.1	The Instantiation	16
4.2	Preferences and Defeats	17
4.2.1	Failure of all ASPIC ⁺ Preferences	17
4.2.2	The Disjoint Elitist Order	18
4.2.3	Mimicking Prioritised Default Logic	20
4.3	Correspondence of Inferences	23
4.3.1	Uniqueness of Stable Extensions	23
4.3.2	A Helpful Distinction in Prioritised Default Logic	25
4.3.3	The Representation Theorem	26
4.4	Summary	30
5	On the Normative Rationality of this Instantiation	31
5.1	Rational Instantiations of ASPIC ⁺	31
5.2	Well-Definedness of this Instantiation	31
5.3	Reasonableness of the Argument Preference	32
5.3.1	Strict Extensions	32
5.3.2	Reasonableness Defined	33
5.3.3	Is the Structure Preference Order Reasonable?	34
5.4	Summary	36
6	Discussion and Conclusion	36
A	Properties of the Disjoint Elitist Order	38
B	Characterising Non-Blocked Defaults	44

1 Introduction

Dung’s argumentation theory [7] has become established as a general framework for non-monotonic reasoning (NMR). Given a set of well-formed formulae (wffs) Δ in some non-monotonic logic (NML), the arguments and attacks defined by Δ instantiate a Dung argumentation framework. Additionally, a preference relation over the defined arguments can be used to determine which attacks succeed as defeats. The justified arguments are then evaluated under various Dung semantics, and the claims of the sceptically justified arguments¹ identify the inferences from the underlying Δ .

¹i.e. the arguments contained in *all* extensions under some semantics.

More formally, given an argumentation framework AF and a wff θ , the *argumentation-defined inference relation* \sim_{AF} over Δ is $\Delta \sim_{AF} \theta$ iff θ is the conclusion of a sceptically justified argument in AF . Indeed, a correspondence has been shown between \sim_{AF} over Δ , and the instantiating logic's non-monotonic inference relation defined directly over Δ . For example, Reiter's default logic (DL) [7], logic programming [7], defeasible logic [9] and Brewka's preferred subtheories [12] have all been endowed with *argumentation semantics*. This in turn allows the application of argument game proof theories [11] to the process of inference, and the generalisation of these dialectical proof theories to distributed NMR amongst computational agents, whereby agents can engage in argumentation-based dialogues, submitting arguments and counter-arguments from their own non-monotonic knowledge bases [1, 10, 14]. Furthermore, argumentative characterisations of NMR make use of principles familiar in everyday reasoning and debate, thus rendering transparent the reasons for accepting a conclusion and allowing for human participation and inspection of the inference process.

One well-studied NML that has not yet been endowed with argumentation semantics is Brewka's *prioritised default logic* (PDL) [3]. PDL is important because it upgrades DL [16] with priorities over defaults². PDL has also been used to represent the (possibly conflicting) beliefs, obligations, intentions and desires (BOID) of agents, and model how these different categories of mental attitudes override each other in order to generate goals and actions that attain those goals [5].

In this note, we endow PDL with argumentation semantics and prove a correspondence between PDL inference and the inference relation defined by the argumentation semantics. We achieve this by appropriately instantiating the ASPIC⁺ framework for structured argumentation [12, 13]. ASPIC⁺ identifies conditions under which logics and preference relations instantiating Dung's frameworks satisfy the rationality postulates of [6]. In Section 2, we review ASPIC⁺ and PDL. In Section 3, we identify an error in the proof of [12, page 376, Proposition 21], which forces us to slightly modify our choice of argument preference relations in Section 2.2. In Section 4, we define a PDL instantiation of ASPIC⁺. This will involve studying preference relations thus far not considered by ASPIC⁺. We then present a representation theorem proving that inferences defined by the argumentation semantics correspond exactly to inferences in PDL. In Section 5, we discuss to what extent is this ASPIC⁺ instantiation normatively rational. Finally, in Section 6, we conclude with future work and some remarks about our approach.

In Appendix A, we prove some properties of a non-ASPIC⁺ argument preference relation, that is relevant to the PDL instantiation. In Appendix B, we prove an intermediate result necessary for the proof of the representation theorem.

²So that, for example, one can account for recent information taking priority over information in the distant past, or that more specific information should take priority over more general information.

2 Background

In the remainder of this paper we make use of the following notation: \mathbb{N} is the set of natural numbers, with $\mathbb{N}^+ := \mathbb{N} - \{0\}$. For a set X its power set is $\mathcal{P}(X)$, the set of its finite subsets is $\mathcal{P}_{\text{fin}}(X)$, so $X \subseteq_{\text{fin}} Y$ iff X is a finite subset of Y , therefore $X \in \mathcal{P}_{\text{fin}}(Y) \Leftrightarrow X \subseteq_{\text{fin}} Y$. Undefined quantities are denoted by $*$, for example $1/0 = *$ in the real numbers. If $\langle P, \leq \rangle$ is a preset (preordered set) then the strict version of the preorder is $a < b \Leftrightarrow [a \leq b, b \not\leq a]$, which is easily shown to be a strict partial order. For two sets A, B their symmetric difference is $A \ominus B := (A - B) \cup (B - A)$.

2.1 Dung's Abstract Argumentation Theory

We now recap the important definitions of [7]. A *(Dung) argumentation framework* is a directed graph $\langle \mathcal{A}, \mathcal{C} \rangle$, where \mathcal{A} is the *set of arguments* and $\mathcal{C} \subseteq \mathcal{A}^2$ is the *conflict relation* over \mathcal{A} . For arguments $A, B \in \mathcal{A}$ we write $\mathcal{C}(A, B) \Leftrightarrow (A, B) \in \mathcal{C} \Leftrightarrow A$ conflicts with B , i.e. A is used as a counterargument against B . Note that \mathcal{C} can denote either an attack relation defined by a set of instantiating formulae, or the defeat relation defined by determining which attacks succeed as defeats.

In what follows let $S \subseteq \mathcal{A}$ be a set of arguments and $A, B \in \mathcal{A}$. S *conflicts with* B iff $(\exists A \in S) \mathcal{C}(A, B)$. S is *conflict-free* (cf) iff $\mathcal{C} \cap S^2 = \emptyset$. S *defends* A iff $(\forall B \in \mathcal{A}) [\mathcal{C}(B, A) \Rightarrow S \text{ conflicts with } B]$. Let $\text{Def}(S) := \{A \in \mathcal{A} \mid S \text{ defends } A\}$. Then, S is an *admissible set* iff S is cf and $S \subseteq \text{Def}(S)$. An admissible set S is:

- a *complete extension* iff $\text{Def}(S) \subseteq S$;
- a *preferred extension* iff S is a \subseteq -maximal complete extension;
- the *grounded extension* iff S is the \subseteq -least complete extension;
- a *stable extension* iff S is complete and conflicts with all arguments in $\mathcal{A} - S$.

$\mathcal{S} := \{\text{complete}, \text{preferred}, \text{grounded}, \text{stable}\}$ is the set of *Dung semantics*. An argument $A \in \mathcal{A}$ is *sceptically justified* under the semantics $s \in \mathcal{S}$ iff A belongs to all s extensions.

2.2 The ASPIC⁺ Framework

Dung's framework provides an intuitive calculus of opposition for determining the justified arguments based on conflict alone. However, it abstracts from the internal logical structure of arguments, the nature of defeats and how they are determined by preferences, and consideration of the conclusions of the arguments. However, these features are referenced when studying whether any given logical instantiation of a framework yields complete extensions that satisfy the rationality postulates of [6]. ASPIC⁺ [12, 13] provides a structured account of abstract argumentation, allowing one to reference the above features, while at

the same time accommodating a wide range of instantiating logics and preference relations. ASPIC⁺ then identifies conditions under which the instantiation (given arguments, attacks and preferences) results in complete extensions that satisfy the rationality postulates of [6]; such instantiations are *normatively rational*.

2.2.1 Construction of Arguments and Knowledge Bases

An (ASPIC⁺) *argumentation system* is $\langle \mathcal{L}, -, \mathcal{R}_s, \mathcal{R}_d, n \rangle$ where \mathcal{L} is a logical language, $- : \mathcal{L} \rightarrow \mathcal{P}(\mathcal{L})$ is the *contrary function* $\theta \mapsto \bar{\theta}$ that identifies when one wff in \mathcal{L} conflicts with another. Let $\theta_1, \dots, \theta_m, \phi \in \mathcal{L}$ be wffs for $m \in \mathbb{N}$, \mathcal{R}_s is the *set of strict inference rules* of the form $(\theta_1, \dots, \theta_m \rightarrow \phi)$, denoting that if $\theta_1, \dots, \theta_m$ are true then ϕ is true no matter what, and \mathcal{R}_d is the *set of defeasible inference rules* of the form $(\theta_1, \dots, \theta_m \Rightarrow \phi)$, denoting that if $\theta_1, \dots, \theta_m$ are true then ϕ is true, unless there are good reasons not to accept ϕ . Finally $n : \mathcal{R}_d \rightarrow \mathcal{L}$ is a *partial function* that assigns a *name* to *some* of the defeasible rules.

For each rule $r \in \mathcal{R}_s \cup \mathcal{R}_d$ we define two functions:

1. The *antecedent map* is

$$\begin{aligned} \text{Ante} : \mathcal{R} &\rightarrow \mathcal{P}_{\text{fin}}(\mathcal{L}) \\ r := (\theta_1, \dots, \theta_m \rightarrow / \Rightarrow \phi) &\mapsto \text{Ante}(r) := \{\theta_1, \dots, \theta_m\}. \end{aligned} \quad (2.1)$$

Note that *Ante* returns a *finite set* of formulae.

2. The *consequent map* is

$$\begin{aligned} \text{Cons} : \mathcal{R} &\rightarrow \mathcal{L} \\ r := (\theta_1, \dots, \theta_m \rightarrow / \Rightarrow \phi) &\mapsto \text{Cons}(r) := \phi. \end{aligned} \quad (2.2)$$

The names of the rules are unique, i.e. if $\text{Ante}(r) = \text{Ante}(r')$ and $\text{Cons}(r) = \text{Cons}(r')$, then $r = r'$. By equality we mean *syntactic* equality with respect to the instantiating logic, e.g. if \mathcal{L} is propositional logic, $\neg(\theta \rightarrow \phi) \neq (\theta \wedge \neg\phi)$. Two rules r, r' are *equal* iff $\text{Ante}(r) = \text{Ante}(r')$ (syntactically) and $\text{Cons}(r) = \text{Cons}(r')$ (syntactically).

An (ASPIC⁺) *knowledge base* is a set $\mathcal{K} := \mathcal{K}_n \cup \mathcal{K}_p \subseteq \mathcal{L}$ where \mathcal{K}_n is the set of *axioms* and \mathcal{K}_p is the set of *ordinary premises*. Intuitively, the knowledge base consists of the premises used in constructing arguments. Note that \mathcal{K} does not have to be a finite set. Given an argumentation system and knowledge base, an (ASPIC⁺) *argument* is defined inductively:

1. (Base) $[\theta]$ is a *singleton (ASPIC⁺) argument* with $\theta \in \mathcal{K}$, conclusion $\text{Conc}([\theta]) := \theta$, premise set $\{\theta\} \subseteq \mathcal{K}$ and top rule $\text{TopRule}([\theta]) := *$.
2. (Inductive, strict) Let $1 \leq i \leq n$ be an index. For each such i let A_i be an ASPIC⁺ argument with conclusion $\text{Conc}(A_i)$ and premise set $\text{Prem}(A_i)$. If $r := (\text{Conc}(A_1), \dots, \text{Conc}(A_n) \rightarrow \phi) \in \mathcal{R}_s$, then $B := [A_1, \dots, A_n \rightarrow \phi]$ is also an ASPIC⁺ argument with conclusion $\text{Conc}(B) = \phi$, premise set $\text{Prem}(B) := \bigcup_{i=1}^n \text{Prem}(A_i) \subseteq \mathcal{K}$ and $\text{TopRule}(B) = r \in \mathcal{R}_s$.

3. (Inductive, defeasible) Let $1 \leq i \leq n$ be an index. For each such i let A_i be an ASPIC⁺ argument with conclusion $Conc(A_i)$ and premise set $Prem(A_i)$. If $r' := (Conc(A_1), \dots, Conc(A_n) \Rightarrow \phi) \in \mathcal{R}_d$, then $C := [A_1, \dots, A_n \Rightarrow \phi]$ is also an ASPIC⁺ argument with conclusion $Conc(C) = \phi$, premise set $Prem(C) := \bigcup_{i=1}^n Prem(A_i) \subseteq \mathcal{K}$ and $TopRule(C) = r' \in \mathcal{R}_d$.

Let \mathcal{A} be the (unique) set of all arguments constructed in this way. It is clear that arguments are finite objects.

2.2.2 Properties of Arguments

A *subargument* B of A is (informally) an argument where $Prem(B) \subseteq Prem(A)$ and $Conc(B)$ appears as an intermediate conclusion of A attained by the application of the exact same rules³. Given an argument A , its *set of subarguments* is $Sub(A) \subseteq \mathcal{A}$ and its *set of proper subarguments* is $Sub(A) - \{A\} \subseteq \mathcal{A}$. We will also write $A \subseteq_{\text{arg}} B \Leftrightarrow A \in Sub(B)$, and $A \subset_{\text{arg}} B \Leftrightarrow A \in Sub(B) - \{B\}$. It is easy to show that \subseteq_{arg} is a preorder and \subset_{arg} is a strict partial order on \mathcal{A} . Informally, two arguments are *equal* iff they are constructed identically in the above manner⁴. Further, a set $S \subseteq \mathcal{A}$ is *subargument closed* iff $(\forall A \in S) Sub(A) \subseteq S$.

An argument $A \in \mathcal{A}$ is *firm* iff $Prem(A) \subseteq \mathcal{K}_n$, i.e. all of its premises are axioms. Further, $DR(A) \subseteq \mathcal{R}_d$ is the set of defeasible rules applied in constructing A . An argument A is *strict* iff $DR(A) = \emptyset$, else A is *defeasible*.

We define the *conclusion map*

$$\begin{aligned} Conc : \mathcal{A} &\rightarrow \mathcal{L} \\ A &\mapsto Conc(A), \end{aligned} \tag{2.3}$$

which takes an argument and returns its conclusion⁵. We can generalise this to *sets* of arguments as follows:

$$\begin{aligned} Conc' : \mathcal{P}(\mathcal{A}) &\rightarrow \mathcal{P}(\mathcal{L}) \\ S &\mapsto Conc'(S) := \bigcup_{A \in S} Conc(A). \end{aligned} \tag{2.4}$$

We will abuse notation and not distinguish between $Conc'$ and $Conc$ if there is no danger of ambiguity. Similarly, one can also define $Prem : \mathcal{A} \rightarrow \mathcal{P}_{\text{fin}}(\mathcal{L})$, $DR : \mathcal{A} \rightarrow \mathcal{P}_{\text{fin}}(\mathcal{R}_d)$, $TopRule : \mathcal{A} \rightarrow \mathcal{R}$ from how arguments are constructed, and generalise their domains from single arguments \mathcal{A} to sets of arguments $\mathcal{P}(\mathcal{A})$. Further, we define, for all $A \in \mathcal{A}$, $Prem_n(A) := Prem(A) \cap \mathcal{K}_n$, $Prem_p(A) := Prem(A) \cap \mathcal{K}_p$, $SR(A) \subseteq \mathcal{R}_s$ is the set of strict rules applied in constructing A .

Notice in the cases of $Prem$ and DR , the codomains are appropriate powersets and not finite powersets, this is because for *arbitrary* sets of arguments, even

³One can formally define subarguments via how arguments are constructed as described in the previous section.

⁴More formally, argument equality can be defined inductively given how arguments are constructed. This will make \subseteq_{arg} into a partial order.

⁵Do not confuse $Conc : \mathcal{A} \rightarrow \mathcal{L}$ with $Cons : \mathcal{R} \rightarrow \mathcal{L}$ (Equation 2.2, page 5).

though each argument is mapped to its finite set of premises or defeasible rules, the set of arguments can be infinite and the union of infinitely many finite sets does not have to be finite.

2.2.3 Attacks

An argument A *attacks* another argument B , denoted as $A \multimap B$, iff at least one of the following hold, where:⁶

1. A is said to *undermine* attack B on the subargument $B' = [\phi]$ iff

$$[\exists \phi \in Prem_p(B)] Conc(A) \in \overline{\phi},$$

i.e. A conflicts with some ordinary premise of B .

2. There is some $B' \subseteq_{\text{arg}} B$ such that for all $i = 1, \dots, n$,

$$[B''_i \subseteq_{\text{arg}} B, B' = [B''_1, \dots, B''_n \Rightarrow \phi]]$$

and $Conc(A) \in \overline{\phi}$. A is then said to *rebut* attack B on the subargument B' .

3. There is some $B' \subseteq_{\text{arg}} B$ such that $r := TopRule(B') \in \mathcal{R}_d$ and $Conc(A) \in \overline{n(r)}$. A is then said to *undercut* attack B on the subargument B' (by arguing against the application of the defeasible rule r in B).

We then abuse notation to define the *attack relation* as $\multimap \subseteq \mathcal{A}^2$ such that $(A, B) \in \multimap \Leftrightarrow A \multimap B$. A set of arguments $S \subseteq \mathcal{A}$ is *attack-conflict-free* (attack-cf) iff $S^2 \cap \multimap = \emptyset$. By transitivity of \subseteq_{arg} , for all $A, B, C \in \mathcal{A}$, if $A \multimap B$ and $B \subseteq_{\text{arg}} C$, then $A \multimap C$.

2.2.4 Preferences and Defeats

A preference relation over arguments is then used to determine which attacks succeed as defeats. We denote the preference $\preceq \subseteq \mathcal{A}^2$ (not necessarily a preorder for now) such that $A \preceq B \Leftrightarrow A$ is not more preferred than B . The strict version is $A \prec B \Leftrightarrow [A \preceq B, B \not\preceq A]$, and equivalence is $A \approx B \Leftrightarrow [A \preceq B, B \preceq A]$. We define a *defeat* as

$$A \hookrightarrow B \Leftrightarrow (\exists B' \subseteq_{\text{arg}} B) [A \multimap B', A \not\preceq B'] . \quad (2.5)$$

That is to say, A defeats B (on B') iff A attacks B on the subargument B' , and B' is not strictly preferred to A . Notice the comparison is made at the subargument B' instead of the whole argument B . We then abuse notation to define the *defeat relation* as $\hookrightarrow \subseteq \mathcal{A}^2$ such that $(A, B) \in \hookrightarrow \Leftrightarrow A \hookrightarrow B$. By transitivity of \subseteq_{arg} , for all $A, B, C \in \mathcal{A}$, if $A \hookrightarrow B$ and $B \subseteq_{\text{arg}} C$, then $A \hookrightarrow C$. A set of arguments $S \subseteq \mathcal{A}$ is *defeat-conflict-free* (defeat-cf) iff $S^2 \cap \hookrightarrow = \emptyset$.

⁶See [13, Section 2] for a further discussion of why attacks are distinguished in this way.

What is the difference between attack-cf and defeat-cf? Clearly, attack-cf implies defeat-cf but the converse is not true. [12] argues that attack-cf should be the correct notion of conflict-freeness to use when invoking Dung semantics, because the presence of attacks denote disagreement between two arguments, and ideally an agent should not accept two arguments that disagree with each other. However, one can also argue that defeat-cf is the correct notion of conflict-freeness to use (e.g. [15]), because an agent can accept two arguments that attack each other, knowing also that the attack does not succeed due to the preference relation. In the upcoming sections, we will prove the stronger result of attack-cf whenever it is possible to, and defeat-cf follows.

In ASPIC⁺, preferences over arguments are calculated from the argument structure through comparing the fallible information (ordinary premises and defeasible rules) they contain. More formally, \mathcal{K}_p and \mathcal{R}_d are endowed with preorders \leq' and \leq'' respectively⁷. This preorder is then lifted to a set-comparison order \sqsubseteq between the (finite) sets of premises or defeasible rules of the arguments, and then finally to \lesssim , following the method in [12, Section 5].

We now recap this lifting of the preorder $<''$ from \mathcal{R}_d to $\mathcal{P}_{\text{fin}}(\mathcal{R}_d)$. We omit comparing premises because in our instantiation we only compare defeasible rules as there are no ordinary premises (Section 4, page 16).

More formally, ASPIC⁺ considers two ordering principles called *democratic* and *elitist* [12, Section 5]⁸, such that for $A, B \in \mathcal{A}$ and $DR(A) \subseteq \mathcal{R}_d$, we define $A \lesssim B$ to be

$$DR(A) \sqsubseteq_{\text{Eli}} DR(B) \quad (2.6)$$

$$\text{or } DR(A) \sqsubseteq_{\text{Dem}} DR(B), \quad (2.7)$$

where, for⁹ $\Gamma, \Gamma' \subseteq_{\text{fin}} \mathcal{R}_d$,

$$\Gamma \sqsubseteq_{\text{Eli}} \Gamma' \Leftrightarrow [\Gamma = \Gamma' \text{ or } \Gamma \triangleleft_{\text{Eli}} \Gamma'] , \quad (2.8)$$

$$\Gamma \triangleleft_{\text{Eli}} \Gamma' \Leftrightarrow (\exists x \in \Gamma) (\forall y \in \Gamma') x <'' y , \quad (2.9)$$

$$\Gamma \sqsubseteq_{\text{Dem}} \Gamma' \Leftrightarrow [\Gamma = \Gamma' \text{ or } \Gamma \triangleleft_{\text{Dem}} \Gamma'] , \text{ and } \quad (2.10)$$

$$\Gamma \triangleleft_{\text{Dem}} \Gamma' \Leftrightarrow (\forall x \in \Gamma) (\exists y \in \Gamma') x <'' y , \quad (2.11)$$

It is easy to show that \lesssim in both cases is a preorder on \mathcal{A} . We define equivalence of arguments as follows:

$$A \approx B \Leftrightarrow DR(A) = DR(B) . \quad (2.12)$$

Note that Equations 2.8 and 2.9 are not exactly the same as [12, page 375, Definition 19]. We will explain this in Section 3 (page 11).

In summary, when comparing two arguments $A, B \in \mathcal{A}$, $A \lesssim B$ iff $[A \approx B$ (Equation 2.12) or $A \prec B]$. In the latter case, $A \prec B \Leftrightarrow DR(A) \triangleleft_{\text{Eli}} DR(B)$, or $A \prec B \Leftrightarrow DR(A) \triangleleft_{\text{Dem}} DR(B)$, depending on which ordering principle is being used.

⁷Where the bigger item is *more preferred*.

⁸See [13, Section 3.5] for a further discussion of both these ordering principles.

⁹It suffices to consider finite sets as arguments are finite.

Given the preference relation \lesssim between arguments, we call the structure $\langle \mathcal{A}, \rightarrow, \lesssim \rangle$ an *ASPIC⁺ SAF* (structured argumentation framework), or *attack graph*. Its corresponding *defeat graph* is $\langle \mathcal{A}, \hookrightarrow \rangle$, where \hookrightarrow is defined in terms of \rightarrow and \lesssim as in Equation 2.5.

2.2.5 Applying Dung’s Semantics

Given $\langle \mathcal{A}, \hookrightarrow \rangle$ one can then evaluate the extensions under Dung’s semantics (Section 2.1, page 4), and thus identify the argumentation defined inferences as follows. Let AS be an argumentation system. The *argumentation-defined inference relation* \vdash_{AS} is $\mathcal{K} \vdash_{AS} \theta$ iff $\theta = \text{Conc}(A)$ where $A \in \mathcal{A}$ is a sceptically justified argument.

2.2.6 Conditions for Normative Rationality

Instantiations of ASPIC⁺ should satisfy some properties to ensure it is normatively rational [6]. Let $\mathcal{E} \subseteq \mathcal{A}$ be a complete extension. Informally, *subargument closure* states that if an argument is in \mathcal{E} , then all its subarguments are in \mathcal{E} . *Closure under strict rules* states that if the conclusions of arguments in \mathcal{E} strictly entail some ϕ , then \mathcal{E} contains an argument concluding ϕ . Finally, *consistency* states that $\text{Conc}(\mathcal{E}) := \bigcup_{A \in \mathcal{E}} \text{Conc}(A)$ is consistent with respect to the instantiating logic. Collectively these are the (*Caminada-Amgoud*) *rationality postulates*.

ASPIC⁺ then identifies sufficient conditions for an instantiation to satisfy these rationality postulates. These are that the instantiation is *well-defined* and that the argument preference ordering \lesssim is *reasonable*. We will say more about these conditions in Section 5 (page 31), where we discuss whether the ASPIC⁺ characterisation of PDL satisfies the rationality postulates.

2.3 Prioritised Default Logic

2.3.1 First Order Logic

In this section we recap PDL [3]. We work in full first order logic (FOL) where the set of first-order formulae is \mathcal{FL} and the set of closed first order formulae¹⁰ is $\mathcal{SL} \subseteq \mathcal{FL}$, with the usual quantifiers and connectives. Given $S \subseteq \mathcal{FL}$, the *deductive closure* (of S) is $\text{Th}(S)$, and given $\theta \in \mathcal{FL}$, the *addition operator* $+$: $\mathcal{P}(\mathcal{FL}) \times \mathcal{FL}$ is defined as $S + \theta := \text{Th}(S \cup \{\theta\})$.

2.3.2 Normal Defaults

A *normal default* is an expression $\frac{\theta:\phi}{\phi}$ where $\theta, \phi \in \mathcal{FL}$ and read “if θ is the case and ϕ is consistent with what we know, then ϕ is the case”¹¹. In this case we call θ the *antecedent* and ϕ the *consequent*. A normal default $\frac{\theta:\phi}{\phi}$ is *closed* iff $\theta, \phi \in \mathcal{SL}$. We will assume all defaults are closed and normal unless stated

¹⁰i.e. first order formulae without free variables

¹¹There are other possible interpretations of normal defaults, see Example 1.

otherwise. Given $S \subseteq \mathcal{SL}$, a default is *active (in S)* iff $[\theta \in S, \phi \notin S, \neg\phi \notin S]$. Intuitively, the first requirement says we need to know the antecedent before applying the default, the second requirement is that the consequent must add new information, and the third requirement ensures that what we infer is consistent with what we know.

2.3.3 Prioritised Default Theories and Extensions

A *finite prioritised default theory* (PDT) is a structure $\langle D, W, \prec \rangle$, where $W \subseteq \mathcal{SL}$ is not necessarily a finite set and $\langle D, \prec \rangle$ is a *finite* strict poset (partially ordered set) of defaults, where $d' \prec d \Leftrightarrow d$ is *more*¹² *prioritised* than d' . Intuitively, W are the known facts and D the defaults that nonmonotonically extend W . We will consider finite PDTs unless otherwise specified.

A PDT's inferences are defined by its extensions. Formally, let $\prec^+ \supseteq \prec$ be a linearisation¹³ of \prec . An *extension (with respect to \prec^+)* is a set $E := \bigcup_{i \in \mathbb{N}} E_i \subseteq \mathcal{SL}$ built inductively as:

$$E_0 := Th(W) \text{ and} \quad (2.13)$$

$$E_{i+1} := \begin{cases} E_i + \phi, & \text{condition 1} \\ E_i, & \text{else} \end{cases} \quad (2.14)$$

where “condition 1” iff “ ϕ is the consequent of the \prec^+ -greatest default d active in E_i ”. Intuitively, one first generates all classical consequences from the facts W , and then iteratively adds the nonmonotonic consequences from the most prioritised default to the least. Notice if W is inconsistent then $E_0 = E = \mathcal{FL}$. We will assume W is consistent unless stated otherwise.

It can be shown that the ascending chain $E_i \subseteq E_{i+1}$ stabilises at some finite $i \in \mathbb{N}$ and that E is consistent provided that W is consistent. E does not have to be unique because there may be more than one distinct linearisation \prec^+ of \prec . We say the PDT $\langle D, W, \prec \rangle$ *sceptically infers* $\theta \in \mathcal{SL}$ iff $\theta \in E$ for all extensions E .

Henceforth, we will refer to a PDT $\langle D, W, \prec \rangle$ where \prec is a strict total order as a *linearised PDT* (LPDT). If \prec is already total then there is only one way to apply the defaults in D (Equation 2.14), hence the extension is unique and all inferences are sceptical. In what follows, we will use \prec^+ to emphasise that the order is total.

One application of PDL is in modelling how an agent reasons with her beliefs, obligations, intentions and desires (BOID).

Example 1. Suppose a research assistant Alice (a) is considering whether she should teach undergraduates. We can model her mental attitudes as a BOID agent's PDT [5] as follows. Define the predicates $R(x) \Leftrightarrow$ “ x is a research assistant”, $A(x) \Leftrightarrow$ “ x is an academic”, and $T(x) \Leftrightarrow$ “ x is teaching

¹² We have defined the order dually to [3] so as to comply with orderings over the ASPIC⁺ defeasible inference rules.

¹³i.e. \prec^+ is a strict total order and hence $\langle D, \prec^+ \rangle$ is a strict toset (totally ordered set).

(undergraduates)”. Alice is a research assistant, so $W = \{R(a)\}$. She believes that research assistants are academics, so her set of beliefs Bel has the default $\frac{R(a):A(a)}{A(a)}$. She does not want to teach and would rather focus on her research, so her set of desires Des include $\frac{R(a):\neg T(a)}{\neg T(a)}$. However, she is obliged to teach, so her set of obligations Obl include $\frac{A(a):T(a)}{T(a)}$. The set of defaults is $D = Bel \cup Des \cup Obl$, and we assume no other defaults are relevant for this example.

In [5], the relative prioritisations of categories of mental attitudes define different agent types. For example, if Alice is a *realistic selfish agent*, the priority (abuse of notation) is $Obl \prec^+ Des \prec^+ Bel$, and therefore the extension is $Th(\{R(a), A(a), \neg T(a)\})$. She thus generates the goal $\neg T(a)$, i.e. she does not teach. However, if she is a *realistic social agent*, the priority (abuse of notation) is $Des \prec^+ Obl \prec^+ Bel$, and therefore she teaches, as $T(a)$ is in the extension.

3 The Corrected ASPIC⁺ Preferences

3.1 Overview of the Problem

In this section, we show that the elitist set comparison relation [12, page 375, Definition 19] is not reasonable inducing and hence cannot guarantee normative rationality. This is because there is an error in the proof of [12, page 376, Proposition 21]. One should use Prakken’s original strict elitist set comparison relation instead [15], which does result in normatively rational ASPIC⁺ instantiations. This is why Equation 2.9 (page 8) is not the same as the original ASPIC⁺ elitist set comparison relation.

3.2 The Elitist Set Comparison Relation is not Reasonable Inducing

Recall that the property of *reasonable inducing* for a given set comparison relation \sqsubseteq is necessary for ASPIC⁺ instantiations to have normatively rational argument preference relations \preceq based on \sqsubseteq , because they preserve Dung’s fundamental lemma [7, page 327, Lemma 10], as discussed in [12, Section 4.2].

Definition 3.1. (From [12, page 376, Definition 22]) Given a preset $\langle P, \leq \rangle$, a set comparison $\sqsubseteq \subseteq [\mathcal{P}_{\text{fin}}(P)]^2$ is *reasonable inducing* iff

1. \sqsubseteq is transitive.
2. For any $\Gamma_0, \Gamma_1, \dots, \Gamma_n \subseteq_{\text{fin}} P$ (for $n \geq 1$), if

$$\bigcup_{i=1}^n \Gamma_i \triangleleft \Gamma_0 \text{ then} \tag{3.1}$$

- (a) $(\exists 1 \leq i \leq n) \Gamma_i \sqsubseteq \Gamma_0$ and
- (b) $(\exists 1 \leq i \leq n) \Gamma_0 \not\sqsubseteq \Gamma_i$.

Recall the following identity for bounded existential quantifiers: for any unary predicate \mathcal{P} and a family of sets A_i indexed by another set I ,

$$\left(\exists x \in \bigcup_{i \in I} A_i \right) \mathcal{P}(x) \Leftrightarrow (\exists i \in I) (\exists x \in A_i) \mathcal{P}(x). \quad (3.2)$$

We now explain why the proof that \preceq_{Eli} is reasonable inducing [12, page 376, proposition 21] is incorrect by locating the error.

Lemma 3.1. *(The following statement, [12, page 376, Proposition 21], may not be true) \preceq_{Eli} is reasonable inducing.*

Proof. (The following proof, from [12, page 390, Proposition 21], is incorrect) We know that \preceq_{Eli} is transitive and satisfies Property 2(a) of reasonable inducing (Definition 3.1, page 11) from Equation 3.2. Assume for contradiction that property 2(b) is false, i.e.

$$\begin{aligned} & (\forall 1 \leq i \leq n) \Gamma_0 \preceq_{Eli} \Gamma_i \\ \Leftrightarrow & (\forall 1 \leq i \leq n) (\exists x \in \Gamma_0) (\forall y \in \Gamma_i) x \leq y \\ \Rightarrow & (\exists x \in \Gamma_0) (\forall y \in \Gamma_1) x \leq y \text{ (by choosing } i = 1), \\ \Leftrightarrow & (\forall y \in \Gamma_1) x_1 \leq y, \end{aligned} \quad (3.3)$$

where $x_1 \in \Gamma_0$ is the witness to \exists . Now from the assumption of strictly less than in the set comparison relation (Equation 3.1, page 11), we have

$$\begin{aligned} \Gamma_0 \not\preceq_{Eli} \bigcup_{i=1}^n \Gamma_i & \Leftrightarrow (\forall x \in \Gamma_0) \left(\exists y \in \bigcup_{i=1}^n \Gamma_i \right) x \not\leq y \\ & \Leftrightarrow (\forall x \in \Gamma_0) (\exists 1 \leq i \leq n) (\exists y \in \Gamma_i) x \not\leq y \text{ by Equation 3.2} \\ & \Rightarrow (\exists 1 \leq i(x_1) \leq n) (\exists y \in \Gamma_{i(x_1)}) x_1 \not\leq y, \end{aligned} \quad (3.4)$$

where in the last step we have instantiated x under \forall to $x_1 \in \Gamma_0$, which is the witness to Equation 3.3.

(INCORRECT STEP) Assume the witness to \exists in Equation 3.4 is 1, i.e. $i(x_1) = 1$. Of course, there is no guarantee that the witness to the first \exists in Equation 3.4 is the same as the instantiation of the first \forall in “ $(\forall 1 \leq i \leq n) \Gamma_0 \preceq_{Eli} \Gamma_i$ ”.

Running with this, we have from Equations 3.3 and 3.4,

$$(\forall y \in \Gamma_1) x_1 \leq y \text{ and } (\exists y \in \Gamma_1) x_1 \not\leq y. \quad (3.5)$$

Therefore, by instantiating the first quantifier \forall to the witness of the second quantifier \exists , calling it $y_0 \in \Gamma_1$, we have

$$x_1 \leq y_0 \text{ and } x_1 \not\leq y_0, \quad (3.6)$$

which is the purported contradiction. Therefore, $(\exists 1 \leq i \leq n) \Gamma_0 \not\preceq_{Eli} \Gamma_i$. \square

An incorrect proof does not mean that the elitist set comparison relation [12, page 375, Definition 19] is not reasonable inducing. We now show that it is not reasonable inducing with the following counterexample.

Lemma 3.2. *The proposition “if $\bigcup_{i=1}^n \Gamma_i \trianglelefteq_{Eli} \Gamma_0$ then $(\exists 1 \leq i \leq n) \Gamma_0 \not\trianglelefteq_{Eli} \Gamma_i$ ” is false, i.e. Property 2(b) of Definition 3.1 fails for \trianglelefteq_{Eli} .*

Proof. The counterexample is as follows: let $\langle P, \leq \rangle$ be an arbitrary preset such that $a, b, c, d \in P$. Let $\Gamma_0 = \{c, d\}$, $\Gamma_1 = \{a\}$, $\Gamma_2 = \{b\}$ so $\Gamma_1 \cup \Gamma_2 = \{a, b\}$. Let \leq be such that $a \approx c$, $a < d$, $d \leq b$ and $c \not\leq b$. Notice

$$\begin{aligned}
\Gamma_1 \cup \Gamma_2 \trianglelefteq_{Eli} \Gamma_0 &\Leftrightarrow [\Gamma_1 \cup \Gamma_2 \trianglelefteq_{Eli} \Gamma_0 \text{ and } \Gamma_0 \not\trianglelefteq_{Eli} \Gamma_1 \cup \Gamma_2] \\
&\Leftrightarrow [\{a, b\} \trianglelefteq_{Eli} \{c, d\} \text{ and } \{c, d\} \not\trianglelefteq_{Eli} \{a, b\}] \\
\{a, b\} \trianglelefteq_{Eli} \{c, d\} &\Leftrightarrow [(a \leq c, a \leq d) \text{ or } (b \leq c, b \leq d)] . \\
&\Leftrightarrow \text{true as } [a \approx c \Rightarrow a \leq c \text{ and } a < d \Rightarrow a \leq d] . \\
\{c, d\} \not\trianglelefteq_{Eli} \{a, b\} &\Leftrightarrow [(c \not\leq a \text{ or } c \not\leq b) \text{ and } (d \not\leq a \text{ or } d \not\leq b)] \\
&\Leftrightarrow \text{true as } [c \not\leq b \text{ and } a < d \Rightarrow d \not\leq a] . \\
\Gamma_0 \trianglelefteq_{Eli} \Gamma_1 &\Leftrightarrow \{c, d\} \trianglelefteq_{Eli} \{a\} \\
&\Leftrightarrow [c \leq a \text{ or } d \leq a] \\
&\Leftrightarrow \text{true as } [a \approx c \Rightarrow c \leq a] . \\
\Gamma_0 \trianglelefteq_{Eli} \Gamma_2 &\Leftrightarrow \{c, d\} \trianglelefteq_{Eli} \{b\} \\
&\Leftrightarrow [c \leq b \text{ or } d \leq b] \\
&\Leftrightarrow \text{true because } d \leq b .
\end{aligned}$$

Therefore, we have found a situation where $\Gamma_1 \cup \Gamma_2 \trianglelefteq_{Eli} \Gamma_0$, $\Gamma_0 \trianglelefteq_{Eli} \Gamma_1$ and $\Gamma_0 \trianglelefteq_{Eli} \Gamma_2$ are all true. \square

Corollary 3.3. \trianglelefteq_{Eli} is not reasonable inducing.

Proof. Immediate from Definition 3.1 (page 11) and Lemma 3.2 (page 13). \square

This failure of the property of being reasonable inducing allows for counterexamples like [8, Example 5.1]. In that example, two defeasible rules can be equivalent under a suitable preorder without being equal. The original elitist order from [12] does allow for defeasible to be equivalent (i.e. just as preferred as each other) without being equal.

3.3 The Strict Elitist Set Comparison is Reasonable Inducing

Consider the strict version of the elitist order, as originally proposed by Prakken in [15, page 109]. We will show that it is reasonable inducing, at the cost of not allowing distinct defeasible rules to be equivalent, i.e. that the only notion of equivalence is equality. Recall that given a preset $\langle P, \leq \rangle$, its strict counterpart preorder is $a < b \Leftrightarrow [a \leq b, b \not\leq a]$, which is a strict partial order.

Definition 3.2. Let $\langle P, \leq \rangle$ be a preset and form its strict poset $\langle P, < \rangle$. Define the *strict elitist set comparison* \triangleleft'_{Eli} on $\mathcal{P}_{\text{fin}}(P)$ as

$$\Gamma \triangleleft'_{Eli} \Gamma' \Leftrightarrow (\exists x \in \Gamma) (\forall y \in \Gamma') x < y. \quad (3.7)$$

Its *non-strict counterpart* is

$$\Gamma \preceq'_{Eli} \Gamma' \Leftrightarrow [\Gamma = \Gamma' \text{ or } \Gamma \triangleleft'_{Eli} \Gamma']. \quad (3.8)$$

Corollary 3.4. \triangleleft'_{Eli} is *irreflexive*.

Proof. Assume for contradiction that $\Gamma \triangleleft'_{Eli} \Gamma$, which is equivalent to

$$(\exists x \in \Gamma) (\forall y \in \Gamma) x < y.$$

Let $x_0 \in \Gamma$ be the witness to \exists , which means $(\forall y \in \Gamma) x_0 < y$, and one can instantiate y to $y = x_0$, which means $x_0 < x_0$ and hence a contradiction. Therefore, \triangleleft'_{Eli} is irreflexive. \square

Lemma 3.5. The *strict elitist set comparison* (Definition 3.2) is *reasonable inducing*.

Proof. Following Definition 3.1, we have:

1. Transitivity:

$$\begin{aligned} & \Gamma \triangleleft'_{Eli} \Gamma' \triangleleft'_{Eli} \Gamma'' \\ \Leftrightarrow & (\exists x \in \Gamma) (\forall y \in \Gamma') x < y \text{ and } (\exists y \in \Gamma') (\forall z \in \Gamma'') y < z \\ \Leftrightarrow & (\forall y \in \Gamma') x_0 < y \text{ and } (\forall z \in \Gamma'') y_0 < z \\ \Rightarrow & x_0 < y_0 \text{ and } (\forall z \in \Gamma'') y_0 < z \\ \Leftrightarrow & (\forall z \in \Gamma'') x_0 < y_0 < z \\ \Rightarrow & (\forall z \in \Gamma'') x_0 < z \\ \Leftrightarrow & (\exists x \in \Gamma) (\forall z \in \Gamma'') x < z \\ \Leftrightarrow & \Gamma \triangleleft'_{Eli} \Gamma'', \end{aligned}$$

where in the third line $x_0 \in \Gamma$ is the witness to the first \exists , and $y_0 \in \Gamma'$ is the witness to the second \exists . Therefore, \preceq'_{Eli} is transitive.

2. Definition 3.1, Property 2(a): we have

$$\begin{aligned} & \bigcup_{i=1}^n \Gamma_i \triangleleft'_{Eli} \Gamma_0 \\ \Leftrightarrow & \left(\exists x \in \bigcup_{i=1}^n \Gamma_i \right) (\forall y \in \Gamma_0) x < y \\ \Leftrightarrow & (\exists 1 \leq i \leq n) (\exists x \in \Gamma_i) (\forall y \in \Gamma_0) x < y \text{ by Equation 3.2} \\ \Leftrightarrow & (\exists 1 \leq i \leq n) \Gamma_i \triangleleft'_{Eli} \Gamma_0 \text{ by Definition 3.2} \\ \Rightarrow & (\exists 1 \leq i \leq n) \Gamma_i \preceq'_{Eli} \Gamma_0. \end{aligned} \quad (3.9)$$

Therefore, \preceq'_{Eli} satisfies the first property.

3. Definition 3.1, Property 2(b): let $1 \leq i_0 \leq n$ be the witness to the first \exists in Equation 3.9, and $x_{i_0} \in \Gamma_{i_0}$ be the witness to the second \exists in Equation 3.9. Equation 3.9:

$$(\forall y \in \Gamma_0) x_{i_0} < y. \quad (3.10)$$

Now assume for contradiction that

$$\begin{aligned} (\forall 1 \leq i \leq n) \Gamma_0 \leq'_{Eli} \Gamma_i &\Rightarrow \Gamma_0 \leq'_{Eli} \Gamma_{i_0} \\ &\Leftrightarrow (\exists x \in \Gamma_0) (\forall y \in \Gamma_{i_0}) x < y \\ &\Leftrightarrow (\forall y \in \Gamma_{i_0}) x_0 < y, \end{aligned} \quad (3.11)$$

where $x_0 \in \Gamma_0$ in Equation 3.11 is the witness to \exists in the previous line. Now instantiate $y \in \Gamma_0$ in Equation 3.10 to x_0 , and instantiate $y \in \Gamma_{i_0}$ in Equation 3.11 to x_{i_0} . Therefore, we have

$$x_{i_0} < x_0 \text{ and } x_0 < x_{i_0}, \quad (3.12)$$

which is a contradiction. Therefore, $(\exists 1 \leq i \leq n) \Gamma_0 \not\leq'_{Eli} \Gamma_i$ and \leq'_{Eli} satisfies the second property.

This means the strict elitist set comparison is reasonable inducing. \square

Example 2. [8, Example 5.1] Let $\mathcal{L} = \{a_i\}_{i=1}^4$ be closed under (syntactic) negation, the contrary function $-$ denote symmetric negation \neg , $\mathcal{K} = \emptyset$, $\mathcal{R}_d = \{(\top \Rightarrow a_i)\}_{i=1}^4$ and

$$\begin{aligned} \mathcal{R}_s = \{ & (a_1, a_2, a_3 \rightarrow \neg a_4), (a_2, a_3, a_4 \rightarrow \neg a_1), \\ & (a_3, a_4, a_1 \rightarrow \neg a_2), (a_4, a_1, a_2 \rightarrow \neg a_3) \} \end{aligned}$$

such that the preorder \lesssim is such that $d_1 \approx d_2$ and $d_3 \approx d_4$ only (reflexivity and transitivity is implicit). This instantiation is well-defined (Section 2.2.6, page 9). The arguments are $A_i := [\top \Rightarrow a_i]$ for $1 \leq i \leq 4$, and

$$\begin{aligned} B_4 &:= [A_1, A_2, A_3 \rightarrow \neg a_4], B_1 := [A_2, A_3, A_4 \rightarrow \neg a_1], \\ B_2 &:= [A_3, A_4, A_1 \rightarrow \neg a_2], B_3 := [A_4, A_1, A_2 \rightarrow \neg a_3]. \end{aligned}$$

The strict elitist set comparison gives

$$\begin{aligned} \{d_1, d_2, d_3\} &\not\leq'_{Eli} \{d_4\}, \{d_2, d_3, d_4\} \not\leq'_{Eli} \{d_1\}, \\ \{d_3, d_4, d_1\} &\not\leq'_{Eli} \{d_2\}, \{d_4, d_1, d_2\} \not\leq'_{Eli} \{d_3\}, \end{aligned}$$

because (e.g.) there is no defeasible rule in $\{d_1, d_2, d_3\}$ that is strictly less than d_4 . If the witness were d_3 , say, then $d_3 \lesssim d_4$, but $d_4 \lesssim d_3$ as well (rather than $d_4 \not\lesssim d_3$), so $d_3 \not\prec d_4$. Therefore, under the strict elitist set comparison, we have $B_i \not\prec A_i$ (here, \prec denotes the argument preference and not the preorder on defeasible rules), hence $B_i \hookrightarrow A_i$ for $1 \leq i \leq 4$. The possible sets of justified arguments are $\{A_1, A_2, A_3, B_4\}$, $\{A_1, A_2, B_3, A_4\}$, $\{A_1, B_2, A_3, A_4\}$ and $\{B_1, A_2, A_3, A_4\}$, whose conclusion sets are consistent.

3.4 Summary

We conclude that Prakken’s elitist set comparison (Definition 3.2) should be used instead of the original elitist set comparison from [12, page 375, Definition 19] in all all future instantiations of ASPIC⁺, if one would like their ASPIC⁺ instantiation to be normatively rational in the sense of [6], and avoid counterexamples similar to that of [8].

4 From ASPIC⁺ to Prioritised Default Logic

We now instantiate ASPIC⁺ to PDL, define a preference relation over arguments, and prove a representation theorem (Theorem 4.5, page 26), which guarantees that the inferences under the argumentation semantics correspond exactly to the inferences in PDL; this is a soundness and completeness result.

4.1 The Instantiation

Let $\langle D, W, \prec^+ \rangle$ be a LPDT¹⁴. The *corresponding* (ASPIC⁺) instantiation is defined as follows:

1. Our arguments are expressed in FOL, so our set of wffs is \mathcal{FL} , although in practice we only consider \mathcal{SL} .
2. The contrary function – *syntactically* defines conflict in terms of classical negation¹⁵ so $(\forall \theta \in \mathcal{FL}) [\neg \theta \in \bar{\theta} \text{ and } \theta \in \neg \bar{\theta}]$.
3. The set of strict rules \mathcal{R}_s characterises inference in first order classical logic. We leave the specific proof theory implicit.
4. The set of defeasible rules \mathcal{R}_d is defined in terms of D as:

$$\mathcal{R}_d := \left\{ (\theta \Rightarrow \phi) \left| \frac{\theta : \phi}{\phi} \in D \right. \right\}, \quad (4.1)$$

with $n \equiv *$. Clearly, there is a bijection¹⁶ f where

$$f : D \rightarrow \mathcal{R}_d : \frac{\theta : \phi}{\phi} \mapsto f \left(\frac{\theta : \phi}{\phi} \right) := (\theta \Rightarrow \phi) \quad (4.2)$$

and we will define the *strict version of the* preorder \leq'' over \mathcal{R}_d as¹⁷

$$(\theta \Rightarrow \phi) <'' (\theta' \Rightarrow \phi') \Leftrightarrow \frac{\theta : \phi}{\phi} \prec^+ \frac{\theta' : \phi'}{\phi'}. \quad (4.3)$$

¹⁴We will discuss why we only consider LPDTs in Section 6 (page 36).

¹⁵For example, $\neg(\theta \wedge \neg\phi)$ is the contrary of $(\theta \wedge \neg\phi)$, but $(\theta \rightarrow \phi)$, where \rightarrow in this case denotes material implication, is not the contrary of $(\theta \wedge \neg\phi)$.

¹⁶Recall from Section 2.2.1 (page 5) that two defeasible rules are *equal* iff they have the same antecedents and consequent *syntactically*.

¹⁷From Footnote 12 (page 10), we do not need to define $<''$ as the order-theoretic dual to \prec^+ , avoiding potential confusion as to which item is more preferred.

We can see that the strict toset $\langle \mathcal{R}_d, <'' \rangle$ is order isomorphic to $\langle D, \prec^+ \rangle$, where the non-strict version of the order \leq'' abbreviates “either $<''$ or $=$ ”. As we are only considering finite D , \mathcal{R}_d is also finite.

5. The set of axiom premises is $\mathcal{K}_n = W$, because we take W to be the set of facts. Furthermore, $\mathcal{K}_p = \emptyset$.

The set \mathcal{A} of ASPIC⁺ arguments are defined as in Section 2.2 (page 4). It is easy to see that all arguments are firm because $\mathcal{K}_p = \emptyset$, and so there are no undermining attacks. As n is undefined, no attack can be an undercut. Therefore, we only have rebut attacks, where $A \rightarrow B$ iff

$$(\exists B', B'' \subseteq_{\text{arg}} B) \ B' = [B'' \Rightarrow \neg \text{Conc}(A)] . \quad (4.4)$$

4.2 Preferences and Defeats

Defeats are defined as in Equation 2.5 (page 7). So given a suitable argument preference \preceq on the arguments \mathcal{A} and attacks defined in the previous section, we can associate an ASPIC⁺ defeat graph $\langle \mathcal{A}, \hookrightarrow \rangle$ to any LPDT $\langle D, W, \prec^+ \rangle$.

How should the argument preference \preceq be defined based on the strict total order $<''$ over \mathcal{R}_d ? We would want to define \preceq such that the extension of the LPDT $\langle D, W, \prec^+ \rangle$ is given by the conclusions of the justified arguments of the defeat graph $\langle \mathcal{A}, \hookrightarrow \rangle$ instantiated by the corresponding ASPIC⁺ instantiation, and is reasonable [12, page 372, Definition 18].

4.2.1 Failure of all ASPIC⁺ Preferences

Unfortunately, *none* of the four ASPIC⁺ argument preferences – democratic weakest link, democratic last link, elitist weakest link, and elitist last link [12, page 375, Definitions 19 to 21] – are suitable because one can devise simple LPDTs where the prioritised default extension does not correspond to the conclusions of the justified arguments. For example, for the elitist weakest link order, \preceq_{Eli} :

Example 3. Consider the LPDT $\langle D, W, \prec^+ \rangle$ where $W = \{a\}$,

$$D = \left\{ d_1 := \frac{a : b}{b}, d_2 := \frac{b : c}{c}, d_3 := \frac{b : \neg c}{\neg c} \right\} \quad (4.5)$$

and $d_1 \prec^+ d_2 \prec^+ d_3$. By Equation 2.14 (page 10), the prioritised default extension is $Th(\{a, b, \neg c\})$, with d_1 applied first, then d_3 , which blocks d_2 .

In the ASPIC⁺ instantiation: $r_1 <'' r_2 <'' r_3$ (where for $i = 1, 2, 3$, $r_i := f(d_i)$ and f is Equation 4.2). The arguments are $A := [[[a] \Rightarrow b] \Rightarrow c]$ and $B := [[[a] \Rightarrow b] \Rightarrow \neg c]$, which rebut each other at their conclusions.

Under the elitist ordering (Equation 2.8, page 8), it is neither the case that $\{r_1, r_2\} \triangleleft_{\text{Eli}} \{r_1, r_3\}$ nor $\{r_1, r_3\} \triangleleft_{\text{Eli}} \{r_1, r_2\}$. As the sets are not equal, we have $A \not\prec B$, $B \not\prec A$ and $A \not\approx B$. This means $A \hookrightarrow B$ and $B \hookrightarrow A$ by Equation 2.5, which means there are two possible stable extensions $\{A\}$ and $\{B\}$ so that

neither argument is sceptically justified, and so $\neg c$ is not an argumentation-defined inference. However $\neg c$ is a PDL inference. Therefore the elitist weakest link ordering cannot be used to calculate \lesssim .

4.2.2 The Disjoint Elitist Order

One can introduce the *disjoint elitist order*, \sqsubseteq_{DEli} , which ignores shared rules when comparing arguments. This is intuitive because when comparing two arguments we should only focus on the fallible information on which the arguments differ. It is defined as

$$\Gamma \sqsubseteq_{DEli} \Gamma' \Leftrightarrow (\exists x \in \Gamma - \Gamma') (\forall y \in \Gamma' - \Gamma) x <'' y, \quad (4.6)$$

with argument equivalence $A \approx B \Leftrightarrow DR(A) = DR(B)$ and \sqsubseteq_{DEli} defined as usual. We call $<''$ the *underlying (strict) total order* of \sqsubseteq_{DEli} .

If we replace Equation 2.8 with \sqsubseteq_{DEli} , then from $r_2 <'' r_3$, it is easy to see that in Example 3, $A \prec B$, $B \not\prec A$, and so $A \not\prec B$ and $B \hookrightarrow A$, and hence there is only a single stable extension containing the now sceptically justified argument B with conclusion $\neg c$. This at least repairs the correspondence in Example 3.

The disjoint elitist order also satisfies a very intuitive property:

Lemma 4.1. $(\forall A, B \in \mathcal{A}) [DR(A) \subseteq DR(B) \Rightarrow B \lesssim A]$.

Proof. If $DR(B) = DR(A)$ then $B \approx A$, so $B \lesssim A$. If $DR(A) \subset DR(B)$, then $DR(A) - DR(B) = \emptyset$, which means $B \prec A$ is vacuously true from Equation 4.6 so $B \lesssim A$ follows. \square

Formally, this result states that \sqsubseteq_{DEli} extends the superset relation on $\mathcal{P}_{fin}(\mathcal{R}_d)$. Intuitively, this means that the *more* defeasible rules your argument contains, the *less* preferred it will become. It is rational for agents to prefer more certainty than less, with all else being equal¹⁸. The extreme case is that arguments with *no* defeasible rules, i.e. strict arguments, are most preferred.

The above two intuitions of (1) ignoring shared elements of the sets being compared and (2) the order extends the superset relation have been considered in a different context and for a different order by [4].

The disjoint elitist order also satisfies the following property: if $\langle \mathcal{R}_d, <'' \rangle$ is a strict toset, then $\langle \mathcal{P}_{fin}(\mathcal{R}_d), \sqsubseteq_{DEli} \rangle$ is also a strict toset. Further, \emptyset is the \sqsubseteq_{DEli} -greatest element in $\mathcal{P}_{fin}(\mathcal{R}_d)$, and \mathcal{R}_d (if finite) is the \sqsubseteq_{DEli} -least element. See Appendix A (page 38) for proofs of these properties.

Unfortunately, despite these desirable properties, there is a counterexample which shows the disjoint elitist order cannot be used to provide the correspondence with PDL.

¹⁸More precisely, if the agent has two arguments, A and B , such that A is less certain than B because A uses more defeasible rules, then with all else being equal, the agent *should* prefer B over A . However, we do not demand that the agent must first seek complete certainty in the sense of Descartes (especially given limited knowledge and cognitive resources) prior to proposing an argument in a dialogue.

Example 4. Consider the LPDT $\langle D, W, \prec^+ \rangle$ where $D = \{d_1, d_2, d_3, d_4, d_5\}$, $W = \emptyset$ and

$$\begin{aligned} d_1 &:= \frac{\top : a_1}{a_1}, \quad d_4 := \frac{a_3 : a_4}{a_4}, \quad d_3 := \frac{\top : a_3}{a_3}, \\ d_2 &:= \frac{a_1 : a_2}{a_2}, \quad d_5 := \frac{a_1 : \neg(a_2 \wedge a_4)}{\neg(a_2 \wedge a_4)}, \end{aligned}$$

such that $d_1 \prec^+ d_4 \prec^+ d_3 \prec^+ d_2 \prec^+ d_5$. Our PDE is constructed in the usual manner starting from $E_0 = Th(\emptyset)$. By Equation 2.14 (page 10), the order of the application of the defaults is d_3, d_4, d_1, d_5 , with d_2 blocked.

$$E_1 = E_0 + a_3, \quad E_2 = E_1 + a_4, \quad E_3 = E_2 + a_1, \quad E_4 = E_3 + \neg(a_2 \wedge a_4), \quad (4.7)$$

and $E_k = E_4$ for all $k \geq 5$. The default d_2 is blocked because $\neg(a_2 \wedge a_4) \equiv (\neg a_2 \vee \neg a_4)$, and with a_4 (from d_4), we have $\neg a_2$, which blocks d_2 . The unique PDE from this LPDT is

$$E = Th(\{a_1, a_3, a_4, \neg(a_2 \wedge a_4)\}) = Th(\{a_1, \neg a_2, a_3, a_4\}). \quad (4.8)$$

Now consider the corresponding arguments from our instantiation. We have the defeasible rules¹⁹

$$r_1 <'' r_4 <'' r_3 <'' r_2 <'' r_5. \quad (4.9)$$

The relevant arguments and sets of defeasible rules are

$$A := [[\top \Rightarrow a_1] \Rightarrow a_2], \quad DR(A) = \{r_1, r_2\} \quad (4.10)$$

$$B := [[\top \Rightarrow a_3] \Rightarrow a_4], \quad DR(B) = \{r_3, r_4\} \quad (4.11)$$

$$C := [[\top \Rightarrow a_1] \Rightarrow \neg(a_2 \wedge a_4)], \quad DR(C) = \{r_1, r_5\}, \quad (4.12)$$

$$D := [B, C \rightarrow \neg a_2], \quad DR(D) = \{r_1, r_3, r_4, r_5\}, \quad (4.13)$$

We illustrate these arguments in Figure 4.1.

The stable extension is $\{D, B, C, [\top \Rightarrow a_3], [\top \Rightarrow a_1]\}$ and all strict extensions thereof²⁰. The conclusion set does correspond to Equation 4.8. However, this would require $D \hookrightarrow A$, which means, by Equation 2.5 (page 7), $D \multimap A$ and $D \not\prec A$. Clearly, $D \multimap A$ on A . However, $D \not\prec A$ means, under the disjoint elitist order,

$$\begin{aligned} D \not\prec A &\Leftrightarrow \text{not } DR(D) \triangleleft_{DEli} DR(A) \\ &\Leftrightarrow \text{not } (\exists x \in DR(D) - DR(A)) (\forall y \in DR(A) - DR(D)) x <'' y \\ &\Leftrightarrow \text{not } (\exists x \in \{r_1, r_3, r_4, r_5\}) (\forall y \in \{r_2\}) x <'' y \\ &\Leftrightarrow \text{not } (\exists x \in \{r_1, r_3, r_4, r_5\}) x <'' r_2 \\ &\Leftrightarrow \text{not } (r_1 <'' r_2 \text{ or } r_3 <'' r_2 \text{ or } r_4 <'' r_2 \text{ or } r_5 <'' r_2) \\ &\Leftrightarrow r_2 <'' r_1, r_2 <'' r_3, r_2 <'' r_4, r_2 <'' r_5. \end{aligned} \quad (4.14)$$

¹⁹Where, similar to Example 3, r_i corresponds to d_i via Equation 4.2 (page 16).

²⁰Informally, in ASPIC⁺, for $S \subseteq \mathcal{A}$ the strict extension of S is the smallest set containing S extended with all strict and firm arguments, and all possible applications of strict rules to those arguments. This becomes the deductive closure when ASPIC⁺ is instantiated into classical logic. See [12, page 370, Definition 17] for more details.

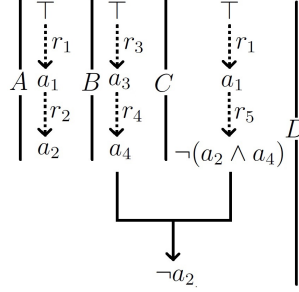


Figure 4.1: The arguments in Example 4. The dashed arrows denote defeasible rules and the solid arrows denote strict rules. Arguments A , B and C are clearly indicated by the label on the line to the left of the constituted argument. Argument D is constructed from arguments B and C , and the strict rule $(a_4, \neg(a_2 \wedge a_4) \rightarrow \neg a_2) \in \mathcal{R}_s$.

From Equation 4.9, it is not the case that $r_2 <'' r_1, r_3, r_4$, so we conclude $D < A$. Therefore, argumentation does not generate the corresponding stable extension to Equation 4.8.

4.2.3 Mimicking Prioritised Default Logic

Despite not being suitable for our desired correspondence between ASPIC⁺ and PDL, the disjoint elitist order does capture one important intuition. When comparing two arguments $A, B \in \mathcal{A}$, we compare them at their defeasible rules, and whichever argument has the $<''$ -least rule in the set $DR(A) \ominus DR(B)$ is the less preferred argument. However, comparing sets of defeasible rules does not take the *structure* of arguments into account, i.e. in terms of which rules in the construction of arguments could be applied earlier, and which could be applied later. We now transform $<''$ into a new order, $<_{SP}$, called the *structure preference* order, such that it gives the correct argument preference for the correspondence.

Let $\langle \mathcal{R}_d, <'' \rangle$ be given. Given $R \subseteq \mathcal{R}_d$, we define the set $Args(R) \subseteq \mathcal{A}$ to be the set of arguments such that $A \in Args(R) \Leftrightarrow DR(A) \subseteq R$. We call this set *the set of arguments freely constructed with defeasible rules in R* . It can be easily shown that $Args$ is \subseteq -monotonic in R and that the assignment $R \mapsto Args(R)$ is functional. Clearly, $Args(\emptyset)$ is the set of all strict arguments, and $Args(\mathcal{R}_d) = \mathcal{A}$. Further, $Args(R)$ is *subargument-closed*, i.e.

Corollary 4.2. *If $A \in Args(R)$ and $B \subseteq_{arg} A$, then $B \in Args(R)$, for any $R \subseteq \mathcal{R}_d$.*

Proof. It is easy to show that if $A \subseteq_{arg} B$, then $DR(A) \subseteq DR(B)$. Therefore, if $A \in Args(R)$, then $DR(A) \subseteq R$, and hence $DR(B) \subseteq R$. Therefore, $B \in Args(R)$. \square

For $R \subseteq \mathcal{R}_d$ let $\max_{<''} R \subseteq R$ denote the *set* of all $<''$ -maximal elements of R . As $<''$ is a (strict) total order and \mathcal{R}_d is finite, this is a singleton set. For $r \in \mathcal{R}_d$ recall the *Ante* map (Equation 2.1, page 5). Note that we are considering defeasible rules with one antecedent, so $\text{Ante}(r)$ is a singleton set. For $S \subseteq \mathcal{A}$ recall the *Conc* map (Equation 2.4, page 6).

Consider ordering the rules in \mathcal{R}_d as follows: for $1 \leq i \leq |\mathcal{R}_d|$, we define the singleton set $\{a_i\} \subseteq \mathcal{R}_d$ to be

$$\max_{<''} \left(\left\{ r \in \mathcal{R}_d \mid \text{Ante}(r) \subseteq \text{Conc} \left(\text{Args} \left(\bigcup_{k=1}^{i-1} \{a_k\} \right) \right) \right\} - \bigcup_{j=1}^{i-1} \{a_j\} \right). \quad (4.15)$$

More concretely,

$$\begin{aligned} \{a_1\} &= \max_{<''} \{r \in \mathcal{R}_d \mid \text{Ante}(r) \subseteq \text{Conc}(\text{Args}(\emptyset))\}, \\ \{a_2\} &= \max_{<''} (\{r \in \mathcal{R}_d \mid \text{Ante}(r) \subseteq \text{Conc}(\text{Args}(\{a_1\}))\} - \{a_1\}), \\ \{a_3\} &= \max_{<''} (\{r \in \mathcal{R}_d \mid \text{Ante}(r) \subseteq \text{Conc}(\text{Args}(\{a_1, a_2\}))\} - \{a_1, a_2\}), \end{aligned}$$

and so on, until this process stops at $\{a_{|\mathcal{R}_d|}\}$ ²¹. The intuition is: a_1 is the most preferred rule whose antecedent is amongst the conclusions of all strict (and firm) arguments, a_2 is the next most preferred rule, whose antecedent is amongst the conclusions of all arguments having at most a_1 as a defeasible rule. Similarly, a_3 is the next most preferred rule, whose antecedent is amongst the conclusions of all arguments having at most a_1 and a_2 as defeasible rules, and so on until all of the rules of \mathcal{R}_d are exhausted. This process orders the rules by how preferred they are under $<''$ and by how earlier they are applicable when constructing the arguments.

We then define (notice the dual order)

$$a_i <_{SP} a_j \Leftrightarrow j < i, \quad (4.16)$$

where $1 \leq i, j \leq |\mathcal{R}_d|$. We define the non-strict order to be $a_i \leq_{SP} a_j \Leftrightarrow [a_i = a_j \text{ or } a_i <_{SP} a_j]$. This makes sense because $i \mapsto a_i$ is bijective between \mathcal{R}_d and $\{1, 2, 3, \dots, |\mathcal{R}_d|\}$. Clearly $<_{SP}$ is a strict total order on \mathcal{R}_d . We call this the *structure preference order* on \mathcal{R}_d , which exists and is unique given $<''$. The corresponding argument preference, \prec_{SP} , is $<_{SP}$ under the disjoint elitist order²², i.e.

$$A \prec_{SP} B \Leftrightarrow (\exists x \in DR(A) - DR(B)) (\forall y \in DR(B) - DR(A)) x <_{SP} y, \quad (4.17)$$

²¹This is just Equation 4.15 with $i = |\mathcal{R}_d|$, which is where the assumption that \mathcal{R}_d is a finite set is crucial.

²²We use the disjoint elitist order instead of the usual elitist order because Example 3 (page 17) shows that the usual elitist order does not give the correspondence.

with \lesssim_{SP} defined as usual. We can also define the corresponding set comparison relation, \triangleleft_{SP} , as, for $\Gamma, \Gamma' \subseteq_{\text{fin}} \mathcal{R}_d$,

$$\Gamma \triangleleft_{SP} \Gamma' \Leftrightarrow (\exists x \in \Gamma - \Gamma') (\forall y \in \Gamma' - \Gamma) x <_{SP} y, \quad (4.18)$$

such that $\Gamma \preceq_{SP} \Gamma' \Leftrightarrow [\Gamma \triangleleft_{SP} \Gamma' \text{ or } \Gamma = \Gamma']$, and

$$A \lesssim_{SP} B \Leftrightarrow DR(A) \preceq_{SP} DR(B). \quad (4.19)$$

The name “structure preference order” refers to the fact that this order takes into account both the preference $<''$ and the structure, i.e. when the rule is first applicable during the construction of arguments. This allows us to imitate how PDL applies defaults when calculating extensions.

Example 5. (Example 4, page 19 continued) We have the following:

$$\begin{aligned} \{a_1\} &= \max_{<''} (\{r \in \mathcal{R}_d \mid \text{Ante}(r) \subseteq \text{Conc}(\text{Args}(\emptyset))\} - \emptyset) \\ &= \max_{<''} \{r \in \mathcal{R}_d \mid \text{Ante}(r) \subseteq \{\top\}\} \\ &= \max_{<''} \{r_1, r_3\} = \{r_3\} \implies a_1 = r_3. \end{aligned}$$

$$\begin{aligned} \{a_2\} &= \max_{<''} (\{r \in \mathcal{R}_d \mid \text{Ante}(r) \subseteq \text{Conc}(\text{Args}(\{r_3\}))\} - \{r_3\}) \\ &= \max_{<''} (\{r_1, r_3, r_4\} - \{r_3\}) \\ &= \max_{<''} \{r_1, r_4\} = \{r_4\} \implies a_2 = r_4. \end{aligned}$$

$$\begin{aligned} \{a_3\} &= \max_{<''} (\{r \in \mathcal{R}_d \mid \text{Ante}(r) \subseteq \text{Conc}(\text{Args}(\{r_3, r_4\}))\} - \{r_3, r_4\}) \\ &= \max_{<''} (\{r_1, r_3, r_4\} - \{r_3, r_4\}) \\ &= \max_{<''} \{r_1\} \implies a_3 = r_1. \end{aligned}$$

$$\begin{aligned} \{a_4\} &= \max_{<''} (\{r \in \mathcal{R}_d \mid \text{Ante}(r) \subseteq \text{Conc}(\text{Args}(\{r_3, r_4, r_1\}))\} - \{r_1, r_3, r_4\}) \\ &= \max_{<''} (\{r_2, r_5, r_1, r_3, r_4\} - \{r_1, r_3, r_4\}) \\ &= \max_{<''} \{r_2, r_5\} = \{r_5\} \implies a_4 = r_5. \end{aligned}$$

$$\{a_5\} = \{r_2\} \implies a_5 = r_2.$$

Therefore, we have

$$a_1 = r_3, a_2 = r_4, a_3 = r_1, a_4 = r_5, a_5 = r_2. \quad (4.20)$$

The structure preference order is

$$r_2 <_{SP} r_5 <_{SP} r_1 <_{SP} r_4 <_{SP} r_3. \quad (4.21)$$

Notice that this is precisely the order of how the corresponding normal defaults are added in PDL, as Equation 4.7 (page 19) shows. It is easy to show that the corresponding stable extension under \prec_{SP} corresponds to the PDL inference, because r_2 is now $<_{SP}$ -least, so $D \not\prec_{SP} A$ by Equation 4.14 (page 19).

However, $<_{SP}$ does not necessarily follow the PDL order of applying defaults as the following example illustrates.

Example 6. Consider the LPDT $\langle D, W, \prec^+ \rangle$ where $W = \{a\}$, $d_1 := \frac{a:\neg a}{\neg a}$ and $d_2 := \frac{\top:b}{b}$, such that $d_2 \prec^+ d_1$. The prioritised default extension is $E = Th(\{a, b\})$, where d_1 is blocked by W , so d_2 is the only default added.

Translating this to argumentation, we have $\mathcal{K}_n = \{a\}$, $r_1 := (a \Rightarrow \neg a)$ and $r_2 := (\top \Rightarrow b)$ where for $i = 1, 2$, $r_i = f(d_i)$, such that $r_2 <'' r_1$. The arguments are $A_0 := [a]$, $A_1 := [A_0 \Rightarrow \neg a]$ and $B := [\top \Rightarrow b]$. Applying Equation 4.15 (page 21), we have $r_2 <_{SP} r_1$, which clearly is not the order of how the corresponding defaults are added in PDL.

Yet the correspondence still holds. Clearly $A_0 \hookrightarrow A_1$ because A_0 is strict, so the stable extension is the strict extension of $\{A_0, B\}$, the conclusion set of which is the extension from PDL.

Example 6 highlights how blocked defaults and defeated arguments are related. Where PDL blocks the application of a given default and hence preventing its conclusion from featuring in the extension, ASPIC⁺ allows for the construction of the argument with the corresponding defeasible rule, but that argument is always defeated by another strictly stronger argument and therefore cannot be in any extension.

4.3 Correspondence of Inferences

In this section we prove that the argument preference \prec_{SP} is the suitable order to give a correspondence between PDL and ASPIC⁺ in all cases. Given an LPDT $\langle D, W, \prec^+ \rangle$, we can construct its defeat graph $\langle \mathcal{A}, \hookrightarrow \rangle$ where the ASPIC⁺ arguments \mathcal{A} are constructed following Section 4.1 (page 16), the attacks \hookrightarrow are rebuts at the conclusions of defeasible rules, and the defeats are as in Equation 2.5 (page 7) with \hookrightarrow and \prec_{SP} . It is always possible to construct $<_{SP}$ and hence \prec_{SP} after translating the LPDT to ASPIC⁺.

4.3.1 Uniqueness of Stable Extensions

In this section we show that the ASPIC⁺ defeat graphs $\langle \mathcal{A}, \hookrightarrow \rangle$ that have been constructed from a LPDT $\langle D, W, \prec^+ \rangle$ (following Section 4.1, page 16) each has a unique stable extension. For any starting preference $<''$ on \mathcal{R}_d we first construct $<_{SP}$ following Equations 4.15 and 4.16 (page 21).

Theorem 4.3. *Let $\langle \mathcal{A}, \hookrightarrow, \prec_{SP} \rangle$ be an ASPIC⁺ attack graph constructed from $\mathcal{L} = \mathcal{FL}$, $-$ is \neg , \mathcal{R}_s the rules of proof of FOL, $\langle \mathcal{R}_d, <'' \rangle$ a finite strict toset of defeasible rules, \prec_{SP} is $<_{SP}$ under the disjoint elitist order, $n \equiv *$ on \mathcal{R}_d ,*

$\mathcal{K}_p = \emptyset$ and $\mathcal{K}_n \subseteq \mathcal{FL}$ is a consistent set of formulae. The defeat graph $\langle \mathcal{A}, \hookrightarrow \rangle$ from this attack graph has a unique stable extension.

Proof. The construction of the unique stable extension mimics how extensions are constructed over an LPDT (Equation 2.14, page 10). Given a set of arguments $S \subseteq \mathcal{A}$ we define, for $r \in \mathcal{R}_d$, $S \oplus r := \text{Args}(\text{DR}(S) \cup \{r\})$, i.e. we close S under all arguments with the addition of a new defeasible rule r . Now consider Algorithm 1. We input the ASPIC⁺ attack graph $\langle \mathcal{A}, \rightarrow, \lesssim_{SP} \rangle$ as described by the hypothesis of the theorem, and the algorithm outputs a set of arguments.

Algorithm 1 Generating a Stable Extension

```

1: function GENERATESTABLEEXTENSION( $\langle \mathcal{A}, \rightarrow, \lesssim_{SP} \rangle$ )
2:    $S \leftarrow \text{Args}(\emptyset)$ 
3:   for  $r \in \mathcal{R}_d$  from  $<_{SP}$ -greatest to  $<_{SP}$ -smallest do
4:     if  $S \oplus r$  is attack-cf then
5:        $S \leftarrow S \oplus r$ 
6:   return  $S$ 

```

The intuition of this algorithm is to create the largest possible set of undefeated arguments, first by including all strict arguments because strict arguments are never defeated (Line 2) and never attack each other because \mathcal{K}_n is consistent. Then, the algorithm includes the defeasible rules from most to least preferred and tests whether the resulting arguments that are constructed by the inclusion of such a defeasible rule leads to an attack (Lines 4–5). As $<_{SP}$ is total, all defeasible rules are considered. This algorithm halts because \mathcal{R}_d is finite.

It is clear from the algorithm that S exists and is unique, because S is a set of freely-constructed arguments (i.e. of the form $\text{Args}(R)$ for some $R \subseteq \mathcal{R}_d$) including as many mutually compatible defeasible rules as possible. We now show that S is a stable extension [2, page 26 Definition 2.2.7].

Attack-cf: This is guaranteed by the consistency of \mathcal{K}_n , so two strict arguments cannot attack each other, and that defeasible rules $r \in \mathcal{R}_d$ are only added if attack-cf is preserved (Lines 4–5). Therefore, S must be attack-cf.

Defeats all other arguments: Let $R \subseteq \mathcal{R}_d$ be the set of all defeasible rules added to S , i.e.

$$R := \text{DR}(S) = \bigcup_{A \in S} \text{DR}(A) \subseteq \mathcal{R}_d. \quad (4.22)$$

Let $B \notin S$ be any argument. As $S = \text{Args}(\text{DR}(S))$, this means there is some rule $r \in \text{DR}(B)$ such that $r \notin R$. The only reason why $r \notin R$ is because if r were added to R , then the resulting S would not be attack-cf, according to Algorithm 1. Let $B' \subseteq_{\text{arg}} B$ be such that $\text{TopRule}(B') = r$, which must exist by the inductive construction of arguments. Let A be the attacker of r , such that²³

²³Note that A is appropriately chosen such that $\text{Conc}(A) = \neg \text{Cons}(r)$ is syntactic equality. This is always possible because \mathcal{R}_s has all rules of proof of FOL. Therefore, if an argument C concludes θ , and we would want it to conclude ϕ , where ϕ and θ are logically equivalent, we can just append the strict rule $(\theta \rightarrow \phi) \in \mathcal{R}_s$ to C to create a new argument D that concludes ϕ .

$\text{Conc}(A) = \neg \text{Cons}(r)$ (Equation 2.2, page 5), so $A \rightarrow B'$ and hence $A \rightarrow B$. There are two possibilities: either r is $<_{SP}$ -greatest or it is not.

Suppose r is $<_{SP}$ -greatest, then $\text{Args}(\emptyset) \oplus r$ is not attack-cf, so $A \in \text{Args}(\emptyset)$. As A is strict, $A \hookrightarrow B'$, and hence $A \hookrightarrow B$.

Now suppose r is not $<_{SP}$ -greatest. Consider the strict up-set of r in \mathcal{R}_d ,

$$T := \{r' \in \mathcal{R}_d \mid r <_{SP} r'\} \neq \emptyset. \quad (4.23)$$

There are two sub-possibilities: either $T \cap R = \emptyset$ or $T \cap R \neq \emptyset$. If the former, then given that adding r to S will create an attack from $A \in \text{Args}(\emptyset)$, we have $A \hookrightarrow B$. If the latter, i.e. $T \cap R \neq \emptyset$, adding r to S means its attacker $A \rightarrow B'$ is in $\text{Args}(T \cap R)$. Either A is strict or not strict (i.e. defeasible). If it is strict, then $A \hookrightarrow B$ as before. If it is not strict, i.e. $\emptyset \neq DR(A) \subseteq T \cap R$, then by definition $(\forall s \in T) r <_{SP} s$. As $DR(A) \subseteq T \cap R \subseteq T$, we must also have $(\forall s \in DR(A)) r <_{SP} s$. Therefore, there is an $r \in DR(B') - DR(A)$ such that for all rules in $DR(A)$, and hence $DR(A) - DR(B')$, $r <_{SP} s$. By Equation 4.17 (page 21), we conclude that $B' \prec_{SP} A$, and hence $A \hookrightarrow B'$. Therefore, by definition of \hookrightarrow and \subseteq_{arg} , $A \hookrightarrow B$.

We conclude that the defeat graphs of these ASPIC⁺ attack graphs have a unique stable extension. \square

4.3.2 A Helpful Distinction in Prioritised Default Logic

In this section we formalise a distinction between defaults in PDL that are blocked because there exists something that disagrees with them, and defaults that are blocked because they do not add any new information.

Let $\langle D, W, \prec \rangle$ be a PDT and $E = \bigcup_{i \in \mathbb{N}} E_i$ one of its extensions generated from the linearisation $\prec^+ \supseteq \prec$. The *set of generating defaults (with respect to \prec^+)*, $GD(\prec^+)$, is defined as

$$\begin{aligned} GD_i(\prec^+) &:= \{d \in D \mid d \text{ is } \prec^+ \text{-greatest active in } E_i\}, \\ GD(\prec^+) &:= \bigcup_{i \in \mathbb{N}} GD_i(\prec^+) \subseteq D. \end{aligned} \quad (4.24)$$

Intuitively, this is the set of defaults applied to calculate E following the order \prec^+ . However, the same E can be generated by distinct total orders \prec^+ .

Example 7. Consider the PDT $\langle \{\frac{a:c}{c}, \frac{b:c}{c}\}, \{a, b\}, \emptyset \rangle$. We have two linearisations $\frac{a:c}{c} \prec_1^+ \frac{b:c}{c}$ and $\frac{b:c}{c} \prec_2^+ \frac{a:c}{c}$. We have $GD(\prec_1^+) = \{\frac{a:c}{c}\}$ and $GD(\prec_2^+) = \{\frac{b:c}{c}\}$, which are not equal, even though both linearisations give the same extension $E = Th(\{a, b, c\})$. But in the case of \prec_1^+ , $\frac{b:c}{c}$ is not active because it adds no new information, rather than that we know $\neg c$ already.

We wish to distinguish between inactive defaults that conflict with something we already know, and inactive defaults that add no new information. We call a default $\frac{\theta:\phi}{\phi}$ *semi-active (in $S \subseteq \mathcal{SL}$)* iff $[\theta \in S, \neg\phi \notin S, \phi \in S]$. The *set of semi-active defaults (with respect to the linearisation \prec^+)* is

$$SAD(\prec^+) := \{d \in D \mid d \text{ is semi-active w.r.t. } \prec^+\}. \quad (4.25)$$

Intuitively, the application of semi-active defaults add no new information. We then define the *set of non-blocked defaults* to be

$$NBD(\prec^+) := GD(\prec^+) \cup SAD(\prec^+) \subseteq D. \quad (4.26)$$

Intuitively, this is precisely the set of defaults which are not blocked by the information that has accumulated in the previous steps. This includes the defaults that we have used to nonmonotonically infer some knowledge, *and* the defaults that do not add any new information.

The set of non-blocked defaults has a more elegant characterisation:

Lemma 4.4. *If \prec^+ generates E , then we have that*

$$NBD(\prec^+) := \left\{ \frac{\theta : \phi}{\phi} \in D \mid \theta \in E, \neg\phi \notin E \right\}. \quad (4.27)$$

Proof. See Appendix B (page 44). \square

Equation 4.27 shows that NBD exists and is unique given an extension E . We may write $NBD(E)$ instead of $NBD(\prec^+)$, or just NBD when E is clear from context. Equation 4.27 adapts Reiter's idea of a *generating default* [16, page 92 Definition 2] to PDL. The set $NBD(E)$ can always be calculated in PDL once E is obtained.

4.3.3 The Representation Theorem

In this section we state and prove the representation theorem, which guarantees that the inferences under the argumentation semantics correspond exactly to the inferences in PDL under \lesssim_{SP} ; this is a soundness and completeness result. More specifically, the theorem relates the (unique) stable extension of $\langle \mathcal{A}, \hookrightarrow \rangle$ with the (unique) prioritised default extension of the corresponding LPDT $\langle D, W, \prec^+ \rangle$.

Theorem 4.5. *Let $\langle \mathcal{A}, \hookrightarrow, \lesssim_{SP} \rangle$ be the attack graph corresponding²⁴ to an LPDT $\langle D, W, \prec^+ \rangle$, with defeat graph $\langle \mathcal{A}, \hookrightarrow \rangle$.*

1. *Let E be the prioritised default extension of $\langle D, W, \prec^+ \rangle$, $NBD(E)$ be the set of non-blocked defaults (Equation 4.27, page 26) and $R := f(NBD(E))$ be the (image) set of corresponding defeasible rules (where f is Equation 4.2, page 16), then $Args(R)$ is the stable extension of $\langle \mathcal{A}, \hookrightarrow \rangle$.*
2. *Let $\mathcal{E} \subseteq \mathcal{A}$ be the unique stable extension of $\langle \mathcal{A}, \hookrightarrow \rangle$ by Theorem 4.3 (page 23), then $Conc(\mathcal{E})$ (Equation 2.4, page 6) is the prioritised default extension of $\langle D, W, \prec^+ \rangle$.*

Proof. We will prove each part separately. For the first statement we will show that the given E is the extension generated from \prec^+ , and $R = f(NBD(E))$ is the set of defeasible rules corresponding to the defaults used in E together

²⁴Recall that we transform \prec^+ to $<''$ using Equation 4.3 (page 16), and then apply Equation 4.16 (page 21) to obtain $<_{SP}$, which gives \lesssim_{SP} as in Equation 4.17.

with the semi-active defaults, the set of arguments $Args(R)$ is a stable extension. For the second statement, we let E be the prioritised default extension of $\langle D, W, \prec^+ \rangle$ and show $Conc(\mathcal{E}) \subseteq E$ and $E \subseteq Conc(\mathcal{E})$.

1. $Args(R)$ is a stable extension:

To show that $Args(R)$ is a stable extension, it is sufficient to show $Args(R)$ is attack-cf and that for all arguments $B \notin Args(R)$, there is some argument $A \in Args(R)$ such that $A \hookrightarrow B$ [2, page 26 Definition 2.2.7].

1.1 $Args(R)$ is attack-cf:

To show that $Args(R)$ is attack-cf we have to show that no two arguments $A, B \in Args(R)$ attack each other. Assume for contradiction that $Args(R)$ is not attack-cf, then $(\exists A, B \in Args(R)) A \rightarrow B$. By definition, $DR(B) \subseteq R = f(NBD(E))$. Suppose $A \rightarrow B$ on $B' \subseteq_{\text{arg}} B$, so $(\exists B'' \subseteq_{\text{arg}} B) B' = [B'' \Rightarrow \neg Conc(A)]$ by Equation 4.4 (page 17). Let $r = TopRule(B')$, so $r = (Conc(B'') \Rightarrow \neg Conc(A))$. Clearly $r \in DR(B)$, and hence $f^{-1}(r) = \frac{Conc(B'') : \neg Conc(A)}{\neg Conc(A)} \in NBD(E)$, and $\neg Conc(A) \in E$.

However, we also have $DR(A) \subseteq R$ as well. Let $\{r_1, \dots, r_n\} \subseteq DR(A)$ be the set of defeasible rules such that $W \cup \{Cons(r_1), \dots, Cons(r_n)\} \models \neg Conc(A)$ ($Cons$ is defined in Equation 2.2, page 5). For $1 \leq i \leq n$ let $d_i := f^{-1}(r_i)$. Clearly the corresponding defaults $d_1, \dots, d_n \in NBE(E)$ and hence $Cons(r_i) \in E$ for $1 \leq i \leq n$. As E is deductively closed and $W \subseteq E$, then $\neg Conc(A) \in E$. Therefore, E is inconsistent - contradiction, because W is consistent. Therefore, $Args(R)$ is attack-cf.

1.2 $Args(R)$ defeats all other arguments:

Now we show that $Args(R)$ defeats all other arguments, by showing that for every argument $B \notin Args(R)$ there exists an argument in $A \in Args(R)$ such that $A \hookrightarrow B$. Let $B \notin Args(R)$ be arbitrary, which means there is some $r \in DR(B)$ such that $r \notin R$. Let $B' \subseteq_{\text{arg}} B$ be such that $TopRule(B') = r$. Let $r = (\theta \Rightarrow \phi)$, so $r \notin R$ means $f^{-1}(r) = \frac{\theta : \phi}{\phi} \notin NBD(E)$. By Equation 4.27 (page 26), this means $\theta \notin E$ or $\neg \phi \in E$. This gives us two possibilities: either $\theta \notin E$, or $\neg \phi \in E$.

1.2.1 The case of $\neg \phi \in E$:

Assume $\neg \phi \in E$, then $(\exists i \in \mathbb{N}) \neg \phi \in E_i$ by Equations 2.13 and 2.14 (page 10). Either $i = 0$ or $i > 0$.

1.2.1.1 The case of $i = 0$:

Suppose $i = 0$, then $W \models \neg \phi$ from Equation 2.13. By compactness, there is some finite $W' \subseteq W$ such that $W' \models \neg \phi$. We can construct an argument

A such that $Prem(A) = W'$ and $Conc(A) = \neg\phi$ as there will be appropriate combinations of strict rules in \mathcal{R}_s , so $A \rightarrow B$. As $DR(A) = \emptyset \subseteq R$, we must have $A \in Args(R)$. Further, as A is strict, $A \hookrightarrow B$ is guaranteed²⁵ by \lesssim_{SP} .

1.2.1.2 The case of $i > 0$:

Now suppose $i > 0$, then $\neg\phi \in E_j$ where $j > 0$ is the witness for i . Let $d_j \in D$ be the default that is \prec^+ -greatest active in the layer E_j , so the set of defaults that conclude $\neg\theta$ (up to the application of deductive rules) is $S := \{d_0, \dots, d_{j-1}\} \subseteq GD_{j-1}(\prec^+) \subseteq NBD(E)$. We can construct an argument A such that $Prem(A) \subseteq W$, $Conc(A) = \neg\phi$ and $DR(A) = f(S)$. Clearly, $DR(A) = f(S) \subseteq f(NBD(E)) = R$ and hence $A \in Args(R)$. It is clear that $A \rightarrow B$, so we need to show $B \not\prec_{SP} A$.

Given that $\neg\phi \in E_j$, it must be the case that $\phi \notin E_j$. Therefore, r is not \prec^+ -greatest active for all extension layers E_0, \dots, E_{j-1} . Suppose for contradiction that there are some rules $s \in DR(A)$ that are \prec_{SP} -smaller than r . Then by Equation 4.15 (page 21), r must be \prec^+ -greatest active at some E_k for $k < j - 1$, which would then result in ϕ in E_{k+1} , therefore preventing $\neg\phi \in E_j$ - contradiction. Therefore, r is \prec_{SP} -smaller than all rules in $DR(A)$. By Equation 4.17, we must have $B \prec_{SP} A$, and hence $A \not\prec_{SP} B$, so $A \hookrightarrow B$. Therefore, for the case of $\neg\phi \in E$, $Args(R)$ defeats all arguments outside it.

1.2.2 The case of $\theta \notin E$:

Now assume $\theta \notin E$. We will show this case is impossible by using the method of infinite descent²⁶.

We start with that $f^{-1}(r) = \frac{\theta:\phi}{\phi}$ and $\theta \notin E$. As $r \in DR(B)$, and $r = TopRule(B')$ (Equation 4.4), there is a $B'' \subset_{arg} B$ that concludes θ . Either B'' is strict or it is not strict. Suppose it is strict, then $Prem(B'') \models \theta$, so by monotonicity $W \models \theta$ because $Prem(B'') \subseteq W$. This means $\theta \in E_0 \subseteq E$ - contradiction. Therefore, B'' cannot be strict.

Furthermore, either $B'' \in Args(R)$ or not. If $B'' \in Args(R)$, then $DR(B'') \subseteq R$, which means $\theta \in E_i \subseteq E$, where i is the level such that all defaults corresponding to $DR(B'')$ have been applied (Equation 2.14) - contradiction. Therefore, $B'' \notin Args(R)$.

This means $DR(B'') \not\subseteq R$, which means there is some rule, $s \in DR(B'')$, such that $f^{-1}(s) \notin NBD(E)$. Suppose $s = \frac{\theta':\phi'}{\phi'}$. There are two possibilities: either $\theta' \notin E$ or $\neg\phi' \in E$. If the latter, then we can construct an argument A' concluding $\neg\phi'$ which defeats B'' as in the case when $\neg\phi \in E$. If the former, we can argue as in the previous paragraph to get a strictly smaller argument $B''' \subset_{arg} B''$ which concludes θ' .

²⁵This is because \lesssim_{SP} is based on the disjoint elitist order, which ranks \emptyset as the greatest element in $\mathcal{P}_{fin}(\mathcal{R}_d)$.

²⁶That is, we argue “backwards” from a given argument $B \in \mathcal{A}$ down to its smallest subarguments (the singletons), and derive a contradiction.

We cannot continue this process forever because arguments are well-founded. Eventually, we must stop at a strict subargument of B'' , which gives a contradiction. Therefore, we cannot have the case $\theta \notin E$. Therefore, this second case is impossible, and the first case means that for every argument $B \notin \text{Args}(R)$ there is some $A \in \text{Args}(R)$ such that $A \hookrightarrow B$. This proves the first statement of the representation theorem.

2. $\text{Conc}(\mathcal{E}) = E$:

We show that $\text{Conc}(\mathcal{E}) \subseteq \mathcal{FL}$ is the prioritised default extension of our LPDT $\langle D, W, \prec^+ \rangle$. We let E be the prioritised default extension of $\langle D, W, \prec^+ \rangle$ and show $E = \text{Conc}(\mathcal{E})$.

2.1 $\text{Conc}(\mathcal{E}) \subseteq E$:

We first show $\text{Conc}(\mathcal{E}) \subseteq E$. Let $\theta \in \text{Conc}(\mathcal{E})$, which means there is some argument $A \in \mathcal{E}$ where $\text{Conc}(A) = \theta$. Either A is strict or it is not.

If A is strict, then as $\text{Prem}(A) \subseteq W$, we must have $W \models \theta$ by monotonicity. Therefore, $W \in E_0 \subseteq E$ by Equation 2.13, and hence $\theta \in E$.

If A is not strict, then for some $k \in \mathbb{N}^+$, $DR(A) := \{d_1, \dots, d_k\}$. None of these defaults give rise to a conflict because \mathcal{E} is a stable extension. Take the smallest $i \in \mathbb{N}$ such that sufficiently many corresponding defeasible rules are applied from $DR(A)$ to conclude θ in E_{i+1} from W . Therefore, $\theta \in E_{i+1} \subseteq E$ and hence $\theta \in E$. Therefore, in either case, $\text{Conc}(\mathcal{E}) \subseteq E$.

2.2 $E \subseteq \text{Conc}(\mathcal{E})$:

We now show $E \subseteq \text{Conc}(\mathcal{E})$. Let $\theta \in E$ so $(\exists i \in \mathbb{N}) \theta \in E_i$ by Equation 2.14 (page 10). We have to show there is some argument $A \in \mathcal{E}$ such that $\text{Conc}(A) = \theta$. Either $i = 0$ or $i > 0$.

2.2.1 *The case of $i = 0$:*

Suppose $i = 0$, which means $\theta \in E_0 \Leftrightarrow W \models \theta$. By compactness, we have some finite $W' \subseteq W$ such that $W' \models \theta$. We can build a strict argument A with $\text{Prem}(A) = W'$ and conclusion θ as \mathcal{R}_s has all the appropriate rules of inference in FOL. Assume for contradiction $A \notin \mathcal{E}$, then there exists some $B \in \mathcal{E}$ defeating A , which is impossible because A is strict. Therefore, $A \in \mathcal{E}$ and $\text{Conc}(A) = \theta$, so $\theta \in \text{Conc}(\mathcal{E})$ by Equation 2.4 (page 6).

2.2.2 *The case of $i > 0$:*

Now suppose $i > 0$. As $\theta \in E_i$, let d_j for $0 \leq j \leq i-1$ be the \prec^+ -greatest active default in E_j . We can use the corresponding defeasible rules $r_j = f(d_j)$ to build an argument A such that $\text{Prem}(A) \subseteq W$, $\text{Conc}(A) = \theta$ and $DR(A) \subseteq \{r_j\}_{j=0}^{i-1}$.

Now we need to show $A \in \mathcal{E}$.

Assume for contradiction that $A \notin \mathcal{E}$, then there is some $B \in \mathcal{E}$ such that $B \hookrightarrow A$. So there is some defeasible rule r in A that is necessary to conclude θ , such that $\text{Conc}(B) = \neg \text{Cons}(r)$. Either B is strict or not.

2.2.2.1 If $B \in \mathcal{E}$ is strict:

Assume that B is strict, then $\text{Conc}(B) \in E_0 \subseteq E_i$, which must conflict with at least one of the rules in $DR(A)$. If this is so, then the corresponding defaults to these rules cannot be \prec^+ -active in the appropriate E_j 's, and hence A cannot be constructed - contradiction. Therefore, B cannot be strict.

2.2.2.2 If $B \in \mathcal{E}$ is not strict:

Assume that B is not strict, then $DR(B) \neq \emptyset$ and, as $B \hookrightarrow A$, there is some $r \in DR(A) - DR(B)$ such that for all $s \in DR(B) - DR(A)$, $r \prec_{SP} s$ by Equation 4.17 (page 21). By Equation 4.15, even if $r \in DR(A) - DR(B)$ can be added to the arguments as a defeasible rule, every single $s \in DR(B) - DR(A)$ is \prec'' -more preferred than r . Therefore, the corresponding defaults in $DR(B) - DR(A)$ are \prec^+ -greatest active in E_j for $j < i$, the application of which would block r from being applied. This contradicts the claim that it is possible to construct A in order to conclude θ . Therefore, B cannot exist.

Therefore, $A \in \mathcal{E}$, and given that $\text{Conc}(A) = \theta$, we have $\theta \in \text{Conc}(\mathcal{E})$ by Equation 2.4. As θ is arbitrary, we conclude $E \subseteq \text{Conc}(\mathcal{E})$ and hence $E = \text{Conc}(\mathcal{E})$. This proves that $\text{Conc}(\mathcal{E})$ is the prioritised default extension of $\langle D, W, \prec^+ \rangle$. \square

The representation theorem allows us to formally interpret the inferences of PDL as the conclusions of justified arguments, and the conclusions of the justified arguments are exactly those of the corresponding PDT. Therefore, by the representation theorem, PDL is sound and complete with respect to its argumentation semantics.

4.4 Summary

In this section, we have provided an instantiation of ASPIC^+ to PDL. We can construct an ASPIC^+ attack graph from a LPDT. The subtlety then is to find a suitable argument preference relation such that it gives a correspondence between the conclusions of the justified arguments, and the extensions of the PDT. We showed that none of the ASPIC^+ orders gives a correspondence, and even the intuitive disjoint elitist order does not give a correspondence either. We then devised the structure-preference order which mimics how defaults are added in PDL when constructing extensions. We then showed that ASPIC^+ defeat graphs that have been constructed by LPDTs have unique stable extensions. The representation theorem states that under the structure-preference order the inferences correspond exactly - this is a soundness and completeness result.

5 On the Normative Rationality of this Instantiation

We have so far instantiated ASPIC⁺ to PDL through an appropriate choice of the underlying logic, defeasible rules and preferences. We have proven that the inferences of this instantiation correspond exactly in Theorem 4.5 (page 26). In this section, we will discuss current work on establishing whether this instantiation is normatively rational²⁷.

5.1 Rational Instantiations of ASPIC⁺

ASPIC⁺ can in principle be instantiated into any concrete argumentation theory, but it is desirable for such instantiations to be normatively rational. For example, a normatively rational instantiation of ASPIC⁺ would guarantee that the conclusions of the ultimately justified arguments are consistent. This idea of normative rationality for structured argumentation frameworks have been formalised by [6]. Let \mathcal{E} denote the set of justified arguments. The rationality postulates informally state [12, Section 4.2]:

1. \mathcal{E} is subargument-closed.
2. $\text{Conc}(\mathcal{E})$ is closed under strict rules.
3. $\text{Conc}(\mathcal{E})$ is consistent.
4. Under \mathcal{R}_s , the closure under strict rules of $\text{Conc}(\mathcal{E})$ is consistent.

ASPIC⁺ provides sufficient conditions for its instantiations to be rational. They are:

1. The argumentation theory (i.e. the argumentation system and the knowledge base) is *well-defined* [12, page 369, Definition 12],
2. and the argument preference relation \succsim is *reasonable* [12, page 372, Definition 18].

5.2 Well-Definedness of this Instantiation

A well-defined classical logic ASPIC⁺ instantiation need only satisfy:

1. *Closure under transposition*: If the rule $(\theta_1, \dots, \theta_n \rightarrow \phi) \in \mathcal{R}_d$ for $n \in \mathbb{N}$, then for all $1 \leq i \leq n$,

$$(\theta_1, \dots, \theta_{i-1}, \neg\phi, \theta_{i+1}, \dots, \theta_n \rightarrow \neg\theta_i) \in \mathcal{R}_s.$$

This is satisfied because \mathcal{R}_s has all the rules of proof of FOL.

²⁷We say “normatively rational” to indicate that the type of rationality we are considering is prescriptive, not descriptive.

2. *Axiom consistency*: This means \mathcal{R}_n is consistent, so we assert that W is consistent²⁸. Given that we are considering LPDTs with consistent W , axiom consistency is satisfied.
3. *Well-formed*: This is a property concerning asymmetric contrary functions, and is vacuously satisfied for instantiations with only a symmetric contrary function, like classical negation \neg .

Therefore, this ASPIC⁺ instantiation into PDL is well-defined.

5.3 Reasonableness of the Argument Preference

One further requirement for an ASPIC⁺ instantiation to be normatively rationality is that the argument preference relation, \preceq , is *reasonable*. For a more detailed discussion of why this is important, see [12, Section 4]. In this subsection, we are concerned with whether the structure preference order, \preceq_{SP} , is reasonable.

5.3.1 Strict Extensions

(From [12, page 370 Definition 17]) Let $S \subseteq_{\text{fin}} \mathcal{A}$. We define a *strict extension* of S to be any *argument* $A \in \mathcal{A}$ that satisfies

$$\begin{aligned} DR(A) &= DR(S), \text{ } Prem_p(A) = Prem_p(S), \\ SR(A) &\supseteq SR(S) \text{ and } Prem_n(A) \supseteq Prem_n(S). \end{aligned} \quad (5.1)$$

The intuition is that given a *finite* set of arguments S , we combine all of these arguments into a bigger argument A only by adding strict rules and axiom premises, leaving the fallible information unchanged. The set S must be finite because the requirement is that *all* arguments in S must be combined into a *single* argument A . We may use the notation S^+ ($\in \mathcal{A}$) instead of A to emphasise that S^+ is a strict extension obtained by extending all arguments of the set S .

Note that given S , S^+ may not exist, nor must it be unique if it does exist. In the former case, there may not be suitable strict rules whose antecedents are in $Conc(S)$ such that S^+ is well-defined, in the sense that S^+ is a single argument. In the latter case, \emptyset^+ exists when $\mathcal{K}_n \neq \emptyset$ but is, by definition, any strict and firm argument (which includes singleton arguments). Clearly, every argument is its own strict extension.

We can define *the set of strict extensions of (a finite set of arguments) S* to be

$$StExt(S) := \{A \in \mathcal{A} \mid \text{Equation 5.1 is true for } A.\} \quad (5.2)$$

Clearly, $\{A\} \subseteq StExt(\{A\})$.

Example 8. Consider an instantiation where $\mathcal{R}_s = \emptyset$, i.e. there are *no strict rules*. Given an arbitrary *finite* set S of arguments, what are the strict extensions in this case?

²⁸The argumentation semantics for PDL will still be valid for an inconsistent W , but normative rationality excludes this case by requiring W to be consistent

1. If $S = \emptyset$, then S^+ is only defined when $\mathcal{K}_n \neq \emptyset$, and S^+ is any of the (strict) singleton arguments. If $\mathcal{K}_n = \emptyset$, then $\emptyset^+ = *$ and hence $StExt(S) = \emptyset$.
2. If $|S| = 1$, say $S = \{A\}$, then $StExt(S) = \{A\}$.
3. If $|S| \geq 2$, then $StExt(S) = \emptyset$ because there are no strict rules to join multiple arguments (or additional axiom premises that may be introduced) in S together into one argument.

Therefore, when $\mathcal{R}_s = \emptyset$, the only case where the strict extension of a set of arguments S is defined is when S is singleton.

5.3.2 Reasonableness Defined

Definition 5.1. (From [12, page 372, Definition 18]) An argument preference relation \prec on \mathcal{A} is *reasonable* iff for all $A, B \in \mathcal{A}$ and $\emptyset \neq S \subseteq_{\text{fin}} \mathcal{A}$,

1. (R1) If A is strict and firm, and B is neither strict nor firm, then $B \prec A$.
2. (R2) If A is strict and firm, then $A \not\prec B$.
3. (R3) If $A \not\prec B$ then $\{A\}^+ \not\prec B$. If $B \not\prec A$ then $B \not\prec \{A\}^+$ (for appropriate strict extensions).
4. (R4) It is *not* the case that

$$(\forall A \in S) (\exists B \in StExt(S - \{A\})) B \prec A. \quad (5.3)$$

The intuition is as follows. (R1) and (R2) state that strict and firm arguments must be maximally preferred. (R3) states that strict extensions do not change the relative preference of arguments. (R4) is an acyclicity condition as illustrated in the following example:

Example 9. Suppose we have a classical logic instantiation of ASPIC⁺ where \mathcal{R}_s have all strict rules and \neg is the only (symmetric) contrary function. Let $\mathcal{R}_d = \{r_1, r_2, r_3\}$ such that $r_1 := (\top \Rightarrow a)$, $r_2 := (\top \Rightarrow b)$ and $r_3 := (\top \Rightarrow \neg(a \wedge b))$. Let $A := [\top \Rightarrow a]$, $B := [\top \Rightarrow b]$ and $C := [\top \Rightarrow \neg(a \wedge b)]$. We also define $F := [A, B \rightarrow (a \wedge b)] \in StExt(\{A, B\})$, $E := [C, A \rightarrow \neg b] \in StExt(\{C, A\})$ and $D := [B, C \rightarrow \neg a] \in StExt(\{B, C\})$. We illustrate these arguments in Figure 5.1.

Clearly, $DR(A) = \{r_1\}$, $DR(B) = \{r_2\}$, $DR(C) = \{r_3\}$, $DR(D) = \{r_2, r_3\}$, $DR(E) = \{r_3, r_1\}$ and $DR(F) = \{r_1, r_2\}$.

Now assume (R4) is false, which means Equation 5.3 is *true*. This means, for $S = \{A, B, C\}$,

$$\begin{aligned} & D \prec A, E \prec B, F \prec C \\ \Leftrightarrow & \{r_2, r_3\} \triangleleft_{DEli} \{r_1\}, \{r_3, r_1\} \triangleleft_{DEli} \{r_2\}, \{r_1, r_2\} \triangleleft_{DEli} \{r_3\} \\ \Leftrightarrow & [r_2 <'' r_1 \text{ or } r_3 <'' r_1], [r_3 <'' r_2 \text{ or } r_1 <'' r_2], [r_1 <'' r_3 \text{ or } r_2 <'' r_3], \end{aligned} \quad (5.4)$$

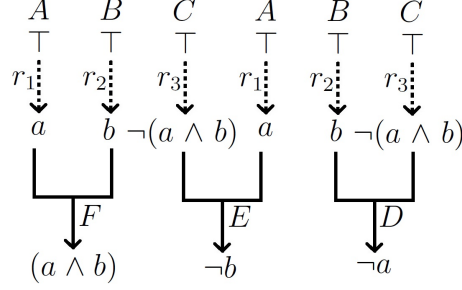


Figure 5.1: The arguments in Example 9.

but no *total* order $<''$ on the set $\{r_1, r_2, r_3\}$ can satisfy any of the above eight conditions (Equation 5.4), as a cycle will always be created. Therefore, \preceq based on the disjoint elitist order satisfies (R4) in this example.

5.3.3 Is the Structure Preference Order Reasonable?

Lemma 5.1. *The structure preference order over arguments, \preceq_{SP} (Equation 4.17, page 21), satisfies (R1) to (R3) in the definition of reasonableness²⁹.*

Proof. We have:

1. (R1) This follows because \emptyset is the greatest element under the disjoint elitist order.
2. (R2) This follows for the same reason as (R1).
3. (R3) This follows because strict extensions do not change the set of defeasible rules.

This shows the result. \square

So to show that \preceq_{SP} is reasonable, we need to show (R4) is true. This is work in progress and we now provide some special cases.

Corollary 5.2. *If $|S| \leq 2$, then \preceq_{SP} is reasonable³⁰.*

Proof. Assume for contradiction that \preceq_{SP} is not reasonable, i.e. Equation 5.3 is true.

If $|S| = 1$, then $S = \{A\}$ (say) so there is a $B \in StExt(S - \{A\}) = StExt(\emptyset)$ such that $B \prec_{SP} A$, so there is a strict (and firm) argument B that is strictly less preferred than A , which contradicts (R2).

²⁹In fact, this result holds for the disjoint elitist order in general, i.e. with any underlying strict total order $<''$. Recall that \preceq_{SP} is just the disjoint elitist order with $<_{SP}$ (Equation 4.16, page 21) as the underlying strict total order.

³⁰This result holds for the disjoint elitist order more generally.

If $|S| = 2$, then $S = \{A, B\}$ (say), so there is a $C \in StExt(\{B\})$ such that $C \prec_{SP} A$, and there is a $D \in StExt(\{A\})$ such that $D \prec_{SP} B$. But by the contrapositive of (R3) this means $A \prec_{SP} B \prec_{SP} A$, which contradicts irreflexivity.

Therefore, for the case of $|S| \leq 2$, \prec_{SP} is reasonable. \square

Lemma 5.3. *If $\mathcal{R}_s = \emptyset$, then \prec_{SP} is reasonable³¹.*

Proof. From Lemma 5.1, it is sufficient to show \prec_{SP} satisfies (R4). Let $\emptyset \neq S \subseteq_{\text{fin}} \mathcal{A}$ be arbitrary. We need to show it is *not* the case that, for any $A \in S$, there is some $B \in StExt(S - \{A\})$ such that $B \prec_{SP} A$. Assume for contradiction that it is true, and let $A \in S$ be arbitrary. We know from Corollary 5.2 (page 34) we need to show this for $|S| > 2$. But from Example 8 (page 32), in the case of no strict rules, $StExt(S - \{A\}) = \emptyset$ for $|S| > 2$, which means Equation 5.3 is false, therefore (R4) holds. \square

Lemma 5.4. *If $\emptyset \neq S \subseteq_{\text{fin}} \mathcal{A}$ is such that for all $A, B \in S$,*

$$A \neq B \implies DR(A) \cap DR(B) = \emptyset, \quad (5.5)$$

then \prec_{SP} is reasonable³².

Proof. Clearly $DR(S) \subseteq_{\text{fin}} \mathcal{R}_d$, because $DR(A) \subseteq_{\text{fin}} \mathcal{R}_d$ for each $A \in \mathcal{A}$ and $DR(S)$ is a union of finite sets. Given the strict toset $\langle \mathcal{R}_d, <'' \rangle$, we also have a *finite* strict toset $\langle DR(S), <'' \rangle$. This has a $<''$ -least element $r_0 \in DR(S)$. By definition, $(\exists A \in S) r_0 \in DR(A)$ (Equation 2.4, page 6). Call the witness to $\exists A_0$, say.

Note that for all *other* arguments in S distinct from A_0 , r_0 would not be amongst their defeasible rules. Now assume for contradiction (R4) is false, so Equation 5.3 is *true*. We instantiate \forall to A_0 and get

$$(\exists B \in StExt(S - \{A_0\})) B \prec_{SP} A_0. \quad (5.6)$$

Let B_0 be the witness to \exists . From the definition of the disjoint elitist order, $B_0 \prec_{SP} A_0$ means that

$$\begin{aligned} & (\exists x \in DR(B_0) - DR(A_0)) (\forall y \in DR(A_0) - DR(B_0)) x <'' y \\ \implies & (\exists x \in DR(B_0)) x <'' r_0, \end{aligned}$$

because $r_0 \in DR(A_0)$ and $r_0 \notin DR(B_0)$. However, as r_0 is the $<''$ -least element of $DR(S)$, there is no element in $DR(B_0)$ that is smaller than r_0 . Therefore, Equation 5.3 is false and hence (R4) is true. \square

Notice that this last result is consistent with Example 9 (page 33). It is still unknown whether \prec_{SP} is reasonable for general sets S where $\emptyset \neq S \subseteq_{\text{fin}} \mathcal{A}$ although we conjecture that it should be given the consistency properties of PDL and that we have shown the representation theorem. This is work in progress.

³¹This result holds for the disjoint elitist order more generally.

³²This result holds for the disjoint elitist order more generally.

5.4 Summary

In this section, we have reviewed the sufficient conditions that an ASPIC^+ instantiation needs to satisfy in order to be normatively rational, which formally means that the rationality postulates of [6] are true. The sufficient conditions are that the instantiation is well-defined, and the argument preference relation is reasonable. It is easy to show that the PDL instantiation is well-defined. We are currently working on showing how the structure preference order, \lesssim_{SP} , is reasonable. We conjecture that it is reasonable, due to the representation theorem.

6 Discussion and Conclusion

In this note we have endowed PDL [3] with argumentation semantics using ASPIC^+ [12]. We did this by instantiating ASPIC^+ to PDL (Section 4, page 16), devising an ASPIC^+ preference order that imitates the procedural construction of extensions in PDL (Section 4.2.3, page 20), and proving the conclusions of the justified arguments correspond exactly to the inferences in PDL (Theorem 4.5, page 26) under this preference. As explained in Section 1, endowing PDL with argumentation semantics allows us to perform inferences in PDL dialectically, in the sense that inference in PDL can formally be viewed as a process of exchanging of arguments and counterarguments [11], until the ultimately justified arguments are found, the conclusions of which are exactly what PDL would conclude. This renders the process of inference in PDL more intuitive, and clarifies the reasons for accepting or rejecting a conclusion.

It is easy to see how Theorem 4.5 generalises the argumentation semantics of preferred subtheories [12, page 381 Theorem 34]. Informally, a default theory is isomorphic to a PDT $\langle D, \emptyset, \prec \rangle$, where D consists of supernormal defaults and \prec is consistent with how the sets of the default theory are ranked. Given a linearisation $\prec^+ \supseteq \prec$, the corresponding preferred subtheory Σ is related to the prioritised default extension by $E = Th(\Sigma)$. The set of arguments with premises from Σ is graph-isomorphic to $Args(f(NBD(E)))$, both of which form a stable extension. Similarly, given the stable extension \mathcal{E} , $Conc(\mathcal{E})$ is the prioritised default extension by Theorem 4.5, and the conclusions of the defeasible rules that feature in \mathcal{E} make up the corresponding preferred subtheory. In future work we will articulate this idea more formally.

There are several issues with the approach we have taken. Firstly, it seems that we have lost generality because we have restricted attention to LPDTs (Section 4.1, page 16). We claim that this does not lose generality because extensions in PDL always presuppose a linearisation \prec^+ of \prec [3], and we have shown that for *any* such linearisation the correspondence between PDL and its argumentation semantics is exact.

Secondly, we have not yet shown that the argument preference relation used, \lesssim_{SP} , is reasonable, so we have not guaranteed normative rationality from the point of view of ASPIC^+ . This is work in progress.

The importance of proving that \lesssim_{SP} is reasonable is that one can use ASPIC⁺ to generalise PDL by abstracting the concepts developed in this note to other situations, not necessarily motivated by PDL. For example, if \lesssim_{SP} is reasonable, then it can be used in a wider range of contexts. Further, ASPIC⁺ can identify argumentation-based inferences assuming only a partial ordering, unlike in PDL. How can multiple partial orderings be related to multiple stable extensions of PDL without explicitly linearising? Also, do the other Dung semantics³³ become relevant? All of this suggests that our argumentative characterisation can be used to generalise PDL, yet if we lift the requirement to linearise, we can no longer guarantee normative rationality, because one can show that the disjoint elitist order is not transitive when the underlying set $\langle \mathcal{R}_d, <'' \rangle$ is a poset instead of a toset³⁴. Future work will consider how to generalise the requirement that the defeasible rules are totally ordered, how in this case one can obtain all PDL extensions via argumentation, and the significance of other types of Dung semantics.

Lastly, the argumentative characterisation of PDL provides for distributed reasoning in the course of deliberation and persuasion dialogues. For example, BOID agents with PDL representations of mental attitudes can now exchange arguments and counterarguments when deliberating about which goals to select, and consequently which actions to pursue. Future work can investigate the precise advantages the argumentation semantics in PDL have over more traditional approaches in such situations.

References

- [1] K. Atkinson, T. Bench-Capon, and P. McBurney. A Dialogue Game Protocol for Multi-Agent Argument over Proposals for Action. *Autonomous Agents and Multi-Agent Systems*, 11(2):153–171, 2005.
- [2] P. Besnard and A. Hunter. *Elements of Argumentation*. The MIT Press, 2008.
- [3] G. Brewka. Adding Priorities and Specificity to Default Logic. In *Logics in Artificial Intelligence*, pages 247–260. Springer, 1994.
- [4] G. Brewka, M. Truszczynski, and S. Woltran. Representing Preferences Among Sets. In *AAAI*, 2010.
- [5] J. Broersen, M. Dastani, J. Hulstijn, and L. van der Torre. Goal Generation in the BOID Architecture. *Cognitive Science Quarterly Journal*, 2(3-4):428–447, 2002.
- [6] M. Caminada and L. Amgoud. On the Evaluation of Argumentation Formalisms. *Artificial Intelligence*, 171(5):286–310, 2007.

³³When the extension is unique, the distinction between the different Dung semantic types is lost.

³⁴See Appendix A, Lemma A.2, page 43.

- [7] P. M. Dung. On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n -Person Games. *Artificial Intelligence*, 77:321–357, 1995.
- [8] P. M. Dung. An Axiomatic Analysis of Structured Argumentation for Prioritised Default Reasoning. In *ECAI2014*, pages 267–272. IOS Press, 2014.
- [9] G. Governatori, M. J. Maher, G. Antoniou, and D. Billington. Argumentation Semantics for Defeasible Logic. *Journal of Logic and Computation*, 14(5):675–702, 2004.
- [10] S. Modgil. An Argumentation Based Semantics for Agent Reasoning. In M. Dastani, A. El Fallah Seghrouchni, J. Leite, and P. Torroni, editors, *Languages, Methodologies and Development Tools for Multi-Agent Systems*, volume 5118 of *Lecture Notes in Computer Science*, pages 37–53. Springer Berlin Heidelberg, 2008.
- [11] S. Modgil and M. Caminada. Proof Theories and Algorithms for Abstract Argumentation Frameworks. In *Argumentation in Artificial Intelligence*, pages 105–129. Springer, 2009.
- [12] S. Modgil and H. Prakken. A General Account of Argumentation with Preferences. *Artificial Intelligence*, 195:361–397, February 2013.
- [13] S. Modgil and H. Prakken. The ASPIC⁺ Framework for Structured Argumentation: a Tutorial. *Argumentation and Computation*, 5(1):31–62, January 2014.
- [14] S. Modgil, F. Toni, and et. al. The Added Value of Argumentation. In S. Ossowski, editor, *Agreement Technologies*, volume 8, pages 357–403. Springer Netherlands, 2013.
- [15] H. Prakken. An Abstract Framework for Argumentation with Structured Arguments. *Argument and Computation*, 1(2):93–124, 2010.
- [16] R. Reiter. A Logic for Default Reasoning. *Artificial Intelligence*, 13:81–132, 1980.

A Properties of the Disjoint Elitist Order

In this section we prove several statements made in Section 4.2.2 (page 18).

Theorem A.1. *If $\langle P, < \rangle$ is a strict toset, then $\langle \mathcal{P}_{\text{fin}}(P), \triangleleft_{\text{DEli}} \rangle$ is also a strict toset.*

Proof. We prove that $\triangleleft_{\text{DEli}}$ is irreflexive, transitive and total over $\mathcal{P}_{\text{fin}}(P)$.

Assume for contradiction $\Gamma \triangleleft_{\text{DEli}} \Gamma$, which is equivalent to, by Equation 4.6 (page 18), $(\exists x \in \emptyset) (\forall y \in \emptyset) x < y$, which is impossible because exists precedes for all. Therefore, $\triangleleft_{\text{DEli}}$ is irreflexive.

To show transitivity, let $n_1, \dots, n_7 \in \mathbb{N}$, such that

$$\begin{aligned} & \{a_1, \dots, a_{n_1}\} \cup \{b_1, \dots, b_{n_2}\} \cup \{c_1, \dots, c_{n_3}\} \cup \{d_1, \dots, d_{n_4}\} \\ & \cup \{e_1, \dots, e_{n_5}\} \cup \{f_1, \dots, f_{n_6}\} \cup \{g_1, \dots, g_{n_7}\} \subseteq P. \end{aligned} \quad (\text{A.1})$$

Each element of the sets are distinct. If $n_i = 0$ then the corresponding set is empty. It is sufficient to consider finite subsets due to $\mathcal{P}_{\text{fin}}(P)$. Let $\Gamma, \Gamma', \Gamma''$ be such that

$$\begin{aligned} \Gamma &= \{a_1, \dots, a_{n_1}\} \cup \{d_1, \dots, d_{n_4}\} \cup \{f_1, \dots, f_{n_6}\} \cup \{g_1, \dots, g_{n_7}\}, \\ \Gamma' &= \{b_1, \dots, b_{n_2}\} \cup \{d_1, \dots, d_{n_4}\} \cup \{e_1, \dots, e_{n_5}\} \cup \{g_1, \dots, g_{n_7}\} \text{ and} \\ \Gamma'' &= \{c_1, \dots, c_{n_3}\} \cup \{e_1, \dots, e_{n_5}\} \cup \{f_1, \dots, f_{n_6}\} \cup \{g_1, \dots, g_{n_7}\}. \end{aligned}$$

We can picture this situation with the the Venn diagram in Figure A.1.

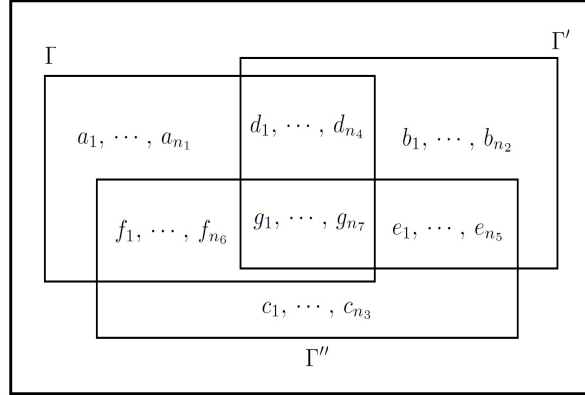


Figure A.1: The solid outer rectangle represents the set P , which may be an infinite set. The three finite sets $\Gamma, \Gamma', \Gamma''$ are the three rectangles within. Each overlapping region has exactly the elements indicated (and nothing more).

Now suppose $<$ permits $\Gamma \triangleleft_{DEli} \Gamma' \triangleleft_{DEli} \Gamma''$, we write this out in terms of

the elements (Equations A.2 and A.3).

$$\begin{aligned}
\Gamma \triangleleft_{DEli} \Gamma' &\Leftrightarrow (\exists x \in \Gamma - \Gamma') (\forall y \in \Gamma' - \Gamma) x < y \\
&\Leftrightarrow (\exists x \in \{a_1, \dots, a_{n_1}\} \cup \{f_1, \dots, f_{n_6}\}) \\
&\quad (\forall y \in \{b_1, \dots, b_{n_2}\} \cup \{e_1, \dots, e_{n_5}\}) x < y \\
&\Leftrightarrow (\exists x \in \{a_1, \dots, a_{n_1}\} \cup \{f_1, \dots, f_{n_6}\}) \\
&\quad \left[\left(\bigwedge_{i=1}^{n_2} x < b_i \right) \wedge \left(\bigwedge_{j=1}^{n_5} x < e_j \right) \right] \\
&\Leftrightarrow \bigvee_{k=1}^{n_1} \left[\left(\bigwedge_{i=1}^{n_2} a_k < b_i \right) \wedge \left(\bigwedge_{j=1}^{n_5} a_k < e_j \right) \right] \\
&\quad \vee \bigvee_{l=1}^{n_6} \left[\left(\bigwedge_{i=1}^{n_2} f_l < b_i \right) \wedge \left(\bigwedge_{j=1}^{n_5} f_l < e_j \right) \right]. \tag{A.2}
\end{aligned}$$

Note that there are $(n_1 + n_6)$ disjuncts in Equation A.2.

$$\begin{aligned}
\Gamma' \triangleleft_{DEli} \Gamma'' &\Leftrightarrow (\exists x \in \Gamma' - \Gamma'') (\forall y \in \Gamma'' - \Gamma') x < y \\
&\Leftrightarrow (\exists x \in \{b_1, \dots, b_{n_2}\} \cup \{d_1, \dots, d_{n_4}\}) \\
&\quad (\forall y \in \{c_1, \dots, c_{n_3}\} \cup \{f_1, \dots, f_{n_6}\}) x < y \\
&\Leftrightarrow (\exists x \in \{b_1, \dots, b_{n_2}\} \cup \{d_1, \dots, d_{n_4}\}) \\
&\quad \left[\left(\bigwedge_{i=1}^{n_3} x < c_i \right) \wedge \left(\bigwedge_{j=1}^{n_6} x < f_j \right) \right] \\
&\Leftrightarrow \bigvee_{k=1}^{n_2} \left[\left(\bigwedge_{i=1}^{n_3} b_k < c_i \right) \wedge \left(\bigwedge_{j=1}^{n_6} b_k < f_j \right) \right] \\
&\quad \vee \bigvee_{l=1}^{n_4} \left[\left(\bigwedge_{i=1}^{n_3} d_l < c_i \right) \wedge \left(\bigwedge_{j=1}^{n_6} d_l < f_j \right) \right]. \tag{A.3}
\end{aligned}$$

Note that there are $(n_2 + n_4)$ disjuncts in Equation A.2.

We need to show that $\Gamma \triangleleft_{DEli} \Gamma''$, i.e.

$$\begin{aligned}
\Gamma \triangleleft_{DEli} \Gamma'' &\Leftrightarrow (\exists x \in \Gamma - \Gamma'') (\forall y \in \Gamma'' - \Gamma) x < y \\
&\Leftrightarrow (\exists x \in \{a_1, \dots, a_{n_1}\} \cup \{d_1, \dots, d_{n_4}\}) \\
&\quad (\forall y \in \{c_1, \dots, c_{n_3}\} \cup \{e_1, \dots, e_{n_5}\}) x < y \\
&\Leftrightarrow (\exists x \in \{a_1, \dots, a_{n_1}\} \cup \{d_1, \dots, d_{n_4}\}) \\
&\quad \left[\left(\bigwedge_{i=1}^{n_3} x < c_i \right) \wedge \left(\bigwedge_{j=1}^{n_5} x < e_j \right) \right] \\
&\Leftrightarrow \bigvee_{k=1}^{n_1} \left[\left(\bigwedge_{i=1}^{n_3} a_k < c_i \right) \wedge \left(\bigwedge_{j=1}^{n_5} a_k < e_j \right) \right] \\
&\quad \vee \bigvee_{l=1}^{n_4} \left[\left(\bigwedge_{i=1}^{n_3} d_l < c_i \right) \wedge \left(\bigwedge_{j=1}^{n_5} d_l < e_j \right) \right]. \tag{A.4}
\end{aligned}$$

To prove Equation A.4, we need to show one of the disjuncts, i.e. for at least one of $1 \leq k \leq n_1$ or $1 \leq l \leq n_4$, we show either

$$\left[\left(\bigwedge_{i=1}^{n_3} a_k < c_i \right) \wedge \left(\bigwedge_{j=1}^{n_5} a_k < e_j \right) \right] \text{ or } \left[\left(\bigwedge_{i=1}^{n_3} d_l < c_i \right) \wedge \left(\bigwedge_{j=1}^{n_5} d_l < e_j \right) \right] \tag{A.5}$$

by establishing all of the conjuncts. Given $\Gamma \triangleleft_{DEli} \Gamma' \triangleleft_{DEli} \Gamma''$, we take the conjunction of Equations A.2 and A.3, making $(n_1 + n_6)(n_2 + n_4)$ disjuncts, which is equivalent to the following expression:

$$\begin{aligned}
&\left\{ \bigvee_{k=1}^{n_1} \left[\left(\bigwedge_{i=1}^{n_2} a_k < b_i \right) \wedge \left(\bigwedge_{j=1}^{n_5} a_k < e_j \right) \right] \vee \bigvee_{l=1}^{n_6} \left[\left(\bigwedge_{i=1}^{n_2} f_l < b_i \right) \wedge \left(\bigwedge_{j=1}^{n_5} f_l < e_j \right) \right] \right\} \\
&\wedge \left\{ \bigvee_{k=1}^{n_2} \left[\left(\bigwedge_{i=1}^{n_3} b_k < c_i \right) \wedge \left(\bigwedge_{j=1}^{n_6} b_k < f_j \right) \right] \vee \bigvee_{l=1}^{n_4} \left[\left(\bigwedge_{i=1}^{n_3} d_l < c_i \right) \wedge \left(\bigwedge_{j=1}^{n_6} d_l < f_j \right) \right] \right\}.
\end{aligned}$$

As \wedge and \vee bi-distribute, we have four cases:

1. For some $1 \leq k \leq n_1$ and $1 \leq k' \leq n_2$, we have

$$\left(\bigwedge_{i=1}^{n_2} a_k < b_i \right) \wedge \left(\bigwedge_{j=1}^{n_5} a_k < e_j \right) \wedge \left(\bigwedge_{i'=1}^{n_3} b_{k'} < c_{i'} \right) \wedge \left(\bigwedge_{j'=1}^{n_6} b_{k'} < f_{j'} \right). \tag{A.6}$$

This means for some $1 \leq k \leq n_1$, we have

$$\left(\bigwedge_{j=1}^{n_5} a_k < e_j \right) \tag{A.7}$$

and from

$$\left(\bigwedge_{i=1}^{n_2} a_k < b_i \right) \wedge \left(\bigwedge_{i'=1}^{n_3} b_{k'} < c_{i'} \right),$$

that $1 \leq k' \leq n_2$, and transitivity of $<$, we infer

$$\left(\bigwedge_{i=1}^{n_3} a_k < c_i \right). \quad (\text{A.8})$$

Equations A.7 and A.8 imply $\Gamma \triangleleft_{DEli} \Gamma''$.

2. For some $1 \leq k \leq n_1$ and $1 \leq l \leq n_4$, we have

$$\left(\bigwedge_{i=1}^{n_2} a_k < b_i \right) \wedge \left(\bigwedge_{j=1}^{n_5} a_k < e_j \right) \wedge \left(\bigwedge_{i'=1}^{n_3} d_l < c_{i'} \right) \wedge \left(\bigwedge_{j'=1}^{n_6} d_l < f_{j'} \right) \quad (\text{A.9})$$

This is the most subtle case of the four, and uses the fact that $<$ is total. The second the third bracketed conjuncts are necessary but not sufficient to lead to $\Gamma \triangleleft_{DEli} \Gamma''$. Let k_0 be the witness to $1 \leq k \leq n_1$ and l_0 be the witness to $1 \leq l \leq n_4$. As $<$ is total, either $a_{k_0} < d_{l_0}$ or $d_{l_0} < a_{k_0}$ (remember all elements are distinct).

- If $a_{k_0} < d_{l_0}$ then $a_{k_0} < c_i$ for all $1 \leq i \leq n_3$. Therefore,

$$\left(\bigwedge_{i=1}^{n_3} a_{k_0} < c_i \right).$$

- If $d_{l_0} < a_{k_0}$ then $d_{l_0} < e_j$ for all $1 \leq j \leq n_5$. Therefore,

$$\left(\bigwedge_{j=1}^{n_5} d_{l_0} < e_j \right).$$

In either case, $\Gamma \triangleleft_{DEli} \Gamma''$.

3. For some $1 \leq l \leq n_6$ and $1 \leq k \leq n_2$, we have

$$\left(\bigwedge_{i=1}^{n_2} f_l < b_i \right) \wedge \left(\bigwedge_{j=1}^{n_5} f_l < e_j \right) \wedge \left(\bigwedge_{i'=1}^{n_3} b_k < c_{i'} \right) \wedge \left(\bigwedge_{j'=1}^{n_6} b_k < f_{j'} \right) \quad (\text{A.10})$$

The irreflexivity of $<$ and the first and last bracketed conjuncts gives a contradiction when you run over all indices, so this case gives a contradiction.

4. For some $1 \leq l \leq n_6$ and $1 \leq l' \leq n_4$, we have

$$\left(\bigwedge_{i=1}^{n_2} f_l < b_i \right) \wedge \left(\bigwedge_{j=1}^{n_5} f_l < e_j \right) \wedge \left(\bigwedge_{i'=1}^{n_3} d_{l'} < c_{i'} \right) \wedge \left(\bigwedge_{j'=1}^{n_6} d_{l'} < f_{j'} \right) \quad (\text{A.11})$$

This case is similar to the first case - we use transitivity to combine the second and last bracketed conjuncts. This infers the second conjunct of Equation A.5, which means $\Gamma \triangleleft_{DEli} \Gamma''$.

Therefore, in all cases, $\Gamma \triangleleft_{DEli} \Gamma''$. This shows \triangleleft_{DEli} is transitive when the underlying order $<$ is total.

Now let $\Gamma, \Gamma' \in \mathcal{P}_{\text{fin}}(P)$ be arbitrary. To show trichotomy, we start by assuming $\Gamma \neq \Gamma'$ and show either $\Gamma \triangleleft_{DEli} \Gamma'$ or $\Gamma' \triangleleft_{DEli} \Gamma$. From Equation 4.6 (page 18), we consider the symmetric difference $\Gamma \ominus \Gamma'$. The set $\langle \Gamma \ominus \Gamma', < \rangle \subseteq \langle P, < \rangle$ is also a finite strict toset. This means there must exist a $<$ -least element $x_0 \in \Gamma \ominus \Gamma'$, say. There are two possibilities:

- If $x_0 \in \Gamma - \Gamma'$, then $\Gamma \triangleleft_{DEli} \Gamma'$.
- If $x_0 \in \Gamma' - \Gamma$, then $\Gamma' \triangleleft_{DEli} \Gamma$.

This establishes trichotomy - so $\langle \mathcal{P}_{\text{fin}}(P), \triangleleft_{DEli} \rangle$ is a strict toset. \square

Lemma A.2. *If $\langle P, < \rangle$ is a strict poset, then \triangleleft_{DEli} is not necessarily transitive over $\mathcal{P}_{\text{fin}}(P)$.*

Proof. We provide the following counterexample: let $P = \{a_0, a_1, a_2, a_3\}$ such that $a_0 < a_2$ and $a_1 < a_3$ and nothing else. This means $a_0 \parallel a_3$ and $a_1 \parallel a_2$. This is a well-defined strict poset. Now let $\Gamma := \{a_0, a_1\}$, $\Gamma' := \{a_1, a_2\}$ and $\Gamma'' := \{a_2, a_3\}$. We can illustrate this in Figure A.2.

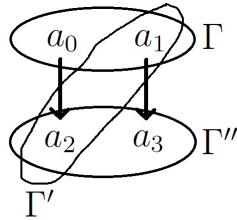


Figure A.2: The Hasse diagram for Lemma A.2. Note that in this document, our Hasse diagrams point the opposite direction, i.e. smaller elements are towards the top, and larger elements are towards the bottom.

Clearly, $\Gamma \triangleleft_{DEli} \Gamma'$ because $a_0 < a_2$, and $\Gamma' \triangleleft_{DEli} \Gamma''$ because $a_1 < a_3$. Now,

$$\begin{aligned} \Gamma \not\triangleleft_{DEli} \Gamma'' &\Leftrightarrow (\forall x \in \Gamma - \Gamma'') (\exists y \in \Gamma'' - \Gamma) x \not\prec y \\ &\Leftrightarrow (\forall x \in \{a_0, a_1\}) (\exists y \in \{a_2, a_3\}) x \not\prec y \\ &\Leftrightarrow (\forall x \in \{a_0, a_1\}) [x \not\prec a_2 \text{ or } x \not\prec a_3] \\ &\Leftrightarrow [a_0 \not\prec a_2 \text{ or } a_0 \not\prec a_3] \text{ and } [a_1 \not\prec a_2 \text{ or } a_1 \not\prec a_3] . \end{aligned}$$

This is true, because $a_0 \not\prec a_3$ and $a_1 \not\prec a_2$ are both true. So this is a situation where $\Gamma \triangleleft_{DEli} \Gamma' \triangleleft_{DEli} \Gamma''$ and $\Gamma \not\triangleleft_{DEli} \Gamma''$. Therefore, \triangleleft_{DEli} is not transitive. \square

Corollary A.3. *If $\langle P, < \rangle$ is a strict toset, then \emptyset is the \triangleleft_{DEli} -greatest element in $\langle \mathcal{P}_{fin}(P), \triangleleft_{DEli} \rangle$.*

Proof. Assume for contradiction that there is some $\Gamma \in \mathcal{P}_{fin}(P)$, $\emptyset \triangleleft_{DEli} \Gamma$, which by Equation 4.6 (page 18) is equivalent to $(\exists x \in \emptyset) (\forall y \in \Gamma) x < y$, which is false due to “ $(\exists x \in \emptyset)$ ”. \square

Corollary A.4. *If $\langle P, < \rangle$ is a strict finite toset, then P is the \triangleleft_{DEli} -least element in $\langle \mathcal{P}_{fin}(P), \triangleleft_{DEli} \rangle$.*

Proof. Assume for contradiction that there is some $\Gamma \in \mathcal{P}_{fin}(P)$, $\Gamma \triangleleft_{DEli} P$. As $\Gamma \subseteq P$, we must have $\Gamma - P = \emptyset$. By Equation 4.6, $\Gamma \triangleleft_{DEli} P$ is equivalent to $(\exists x \in \emptyset) (\forall y \in \Gamma)$, which is false due to “ $(\exists x \in \emptyset)$ ”. \square

B Characterising Non-Blocked Defaults

In this section we prove Lemma 4.4 (page 26). We restate the lemma below for convenience.

Lemma B.1. *Let $\langle D, W, \prec^+ \rangle$ be a LPDT. If \prec^+ generates E , then we have that*

$$NBD(\prec^+) := \left\{ \frac{\theta : \phi}{\phi} \in D \mid \theta \in E, \neg\phi \notin E \right\} . \quad (\text{B.1})$$

Proof. It is sufficient to show Equation 4.26 (with Equations 4.24 and 4.25) is the same as the right hand side of Equation 4.27. Let \prec^+ generate the extension E and, for convenience, we suppress the argument “ \prec^+ ” in the sets for this proof.

(\Rightarrow) Case 1: Assume $d \in SAD$, then $Ante(d) \subseteq S$, $\neg Conc(d) \notin S$ and $Conc(d) \in S$. This implies $Ante(d) \subseteq S$ and $\neg Conc(d) \notin S$. Therefore,

$$d \in \{d' \in D \mid Ante(d') \subseteq E, \neg Conc(d') \notin E\} , \quad (\text{B.2})$$

and hence

$$SAD \subseteq \{d \in D \mid Ante(d) \subseteq E, \neg Conc(d) \notin E\} . \quad (\text{B.3})$$

Case 2: Now assume $d \in GD$, which means

$$\begin{aligned}
&\Leftrightarrow (\exists i \in \mathbb{N}) d \in GD_i \\
&\Leftrightarrow (\exists i \in \mathbb{N}) [Ante(d) \subseteq E_i, Conc(d) \notin E_i, \neg Conc(d) \notin E_i] \\
&\Leftrightarrow Ante(d) \subseteq E_{j_0}, Conc(d) \notin E_{j_0}, \neg Conc(d) \notin E_{j_0} \text{ } j_0 \text{ witness to } i, \quad (B.4) \\
&\Rightarrow Ante(d) \subseteq E_{j_0}, \neg Conc(d) \notin E_{j_0} \\
&\Rightarrow Ante(d) \subseteq E, \neg Conc(d) \notin E_{j_0}.
\end{aligned}$$

Clearly, this means $Ante(d) \subseteq E$.

Now assume for contradiction that $\neg Conc(d) \in E$, which means there is some $i_0 \in \mathbb{N}$ such that $\neg Conc(d) \in E_{i_0}$.

What is the relationship between i_0 and j_0 ? As both are natural numbers, there are three possibilities:

- $i_0 = j_0$: This is impossible as else we will have $\neg Conc(d) \notin E_{i_0}$ and $\neg Conc(d) \in E_{i_0}$.
- $i_0 < j_0$: We have $\neg Conc(d) \notin E_{j_0}$ and $\neg Conc(d) \in E_{i_0}$, which is also impossible because the E_i 's form an ascending chain, so $E_{i_0} \subseteq E_{j_0}$. Therefore, we get $\neg Conc(d) \in E_{j_0}$ and $\neg Conc(d) \notin E_{j_0}$.
- $i_0 > j_0$: We have $\neg Conc(d) \notin E_{j_0}$ and $\neg Conc(d) \in E_{i_0}$. From Equation B.4, we have that d is active in E_{j_0} , hence $Conc(d) \in E_{j_0+1} \subseteq E_{i_0}$, which makes $\neg Conc(d) \in E_{i_0}$ impossible because the E_i 's are consistent.

Therefore, $\neg Conc(d) \notin E$.

So we have $Ante(d) \subseteq E$ and $\neg Conc(d) \notin E$. Therefore, Equation B.2 is true for this case and we have

$$GD \subseteq \{d \in D \mid Ante(d) \subseteq E, \neg Conc(d) \notin E\}. \quad (B.5)$$

We can take the union of Equations B.5 and B.3 to get

$$GD \cup SAD \subseteq \{d \in D \mid Ante(d) \subseteq E, \neg Conc(d) \notin E\}. \quad (B.6)$$

(\Leftarrow) Assume $d \in \{d' \in D \mid Ante(d') \subseteq E, \neg Conc(d') \notin E\}$, which means $Ante(d) \subseteq E$ and $\neg Conc(d) \notin E$. We have, for some $i_0 \in \mathbb{N}$,

$$\begin{aligned}
&\Leftrightarrow \text{Ante}(d) \subseteq E_{i_0}, (\forall j \in \mathbb{N}) \neg \text{Conc}(d) \notin E_j \\
&\Leftrightarrow \text{Ante}(d) \subseteq E_{i_0}, \neg \text{Conc}(d) \notin E_{i_0}, (\forall j \in \mathbb{N} - \{i_0\}) \neg \text{Conc}(d) \notin E_j \\
&\Leftrightarrow (\forall j \in \mathbb{N} - \{i_0\}) \neg \text{Conc}(d) \notin E_j \text{ and} \\
&\quad [(\text{Ante}(d) \subseteq E_{i_0}, \neg \text{Conc}(d) \notin E_{i_0}, \text{Conc}(d) \in E_{i_0}) \text{ or} \\
&\quad (\text{Ante}(d) \subseteq E_{i_0}, \neg \text{Conc}(d) \notin E_{i_0}, \text{Conc}(d) \notin E_{i_0})] \\
&\Leftrightarrow (\forall j \in \mathbb{N} - \{i_0\}) \neg \text{Conc}(d) \notin E_j \text{ and} \\
&\quad [(\text{Ante}(d) \subseteq E_{i_0}, \neg \text{Conc}(d) \notin E_{i_0}, \text{Conc}(d) \in E_{i_0}) \text{ or } d \in GD_{i_0}] \\
&\Rightarrow (\forall j \in \mathbb{N} - \{i_0\}) \neg \text{Conc}(d) \notin E_j \text{ and} \\
&\quad [(\text{Ante}(d) \subseteq E_{i_0}, \neg \text{Conc}(d) \notin E_{i_0}, \text{Conc}(d) \in E_{i_0}) \text{ or } d \in GD] \\
&\Rightarrow d \in GD \text{ or } [\text{Ante}(d) \subseteq E_{i_0}, \neg \text{Conc}(d) \notin E_{i_0}, \text{Conc}(d) \in E_{i_0} \text{ and} \\
&\quad (\forall j \in \mathbb{N} - \{i_0\}) \neg \text{Conc}(d) \notin E_j] \\
&\Rightarrow d \in GD \text{ or } [\text{Ante}(d) \subseteq E \text{ and } (\forall j \in \mathbb{N}) \neg \text{Conc}(d) \notin E_j] \\
&\Rightarrow d \in GD \text{ or } [\text{Ante}(d) \subseteq E \text{ and } \neg \text{Conc}(d) \notin E] , \\
&\Leftrightarrow d \in GD \cup SAD .
\end{aligned}$$

Therefore, we have

$$\{d \in D \mid \text{Ante}(d) \subseteq E, \neg \text{Conc}(d) \notin E\} \subseteq GD \cup SAD . \quad (\text{B.7})$$

The result follows from Equations B.6 and B.7. \square